

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:11 ; Search time 78.64 Seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-74

Sequence: 1 AGMIDAGFKRKITL 14

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

1: PIR.68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	204	2 F64353	dCTP deaminase (EC
2	7	50.0	315	1 D35156	thioredoxin reduct
3	7	50.0	443	1 OOBXPT	hypothetical prote
4	7	50.0	564	2 T26053	hypothetical prote
5	6	42.9	88	2 G81178	conserved hypochet
6	6	42.9	88	2 B81926	hypothetical prote
7	6	42.9	103	2 B64334	hypothetical prote
8	6	42.9	183	1 I64006	hypothetical prote
9	6	42.9	213	2 T17241	hypothetical prote
10	6	42.9	233	2 S74064	hypothetical prote
11	6	42.9	287	1 S72569	hypothetical prote
12	6	42.9	296	2 T48401	probable aldehyde
13	6	42.9	311	2 B70451	histone deacetylase
14	6	42.9	329	2 S04132	phenanthrene dehydro
15	6	42.9	352	2 H70338	photosystem II oxy
16	6	42.9	356	2 E64048	threonine synthase
17	6	42.9	356	2 E64048	nucleotide-binding
18	6	42.9	358	2 D81345	iron utilization p
19	6	42.9	444	2 T46602	cellulobiose/cellotr
20	6	42.9	455	2 D81809	hypothetical prote
21	6	42.9	465	2 T18793	hypothetical prote
22	6	42.9	508	1 A30007	dolichyl-diphospho
23	6	42.9	508	1 ISRTSS	protein disulfide-
24	6	42.9	509	1 ISMSSS	protein disulfide-
25	6	42.9	509	1 A38362	protein disulfide-
26	6	42.9	525	2 T47409	hypothetical prote
27	6	42.9	575	2 T38406	probable flavoprot
28	6	42.9	731	2 T44752	probable malate sy
29	6	42.9	883	2 T37208	hypothetical prote

30	6	42.9	1061	2 T18085	DNA topoisomerase
31	6	42.9	1132	1 OSBP	host specificity p
32	6	42.9	1138	2 D85584	probable tail comp
33	6	42.9	1160	2 A46423	transcription fact
34	6	42.9	1849	2 T14096	guanine nucleotide
35	6	42.9	2104	2 H86127	hypothetical prote
36	5	35.7	19	2 A60894	gamma crystallin I
37	5	35.7	47	2 S69146	gamma-thionin SI-a
38	5	35.7	47	2 S11529	gamma-purothionin
39	5	35.7	47	2 S11530	purothionin gamma
40	5	35.7	47	2 S13849	hordothionin gamma
41	5	35.7	47	2 A58319	gamma-zeathionin I
42	5	35.7	47	2 S69144	gamma-thionin SI-a
43	5	35.7	48	2 S13863	alpha-amylose inh1
44	5	35.7	53	2 I60384	gene T1 protein -
45	5	35.7	92	2 I39925	transcription regu

ALIGNMENTS

RESULT 1
F64353
dCTP deaminase (EC 3.5.4.13) M10430 [similarity] - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: F64353
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
R:Bult, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rsen, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; M10430; M10430; M10430; M10430; M10430; M10430; M10430
A:Accession: F64353
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-204 <BDL>
A:Cross-references: GB:U67494; GB:L77117; NID:g1591120; PIDN:AB98415.1; PID:g1591133
C:Genetics:
A:Map position: REV386963-386349
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGMIDAGFKRKITL 14
Db 131 AGMIDAGFKRKITL 144

RESULT 2
D35156
thioredoxin reductase (NADPH) (EC 1.6.4.5) - Eubacterium acidaminophilum
N:Alternate names: dihydrolipoamide dehydrogenase [misidentification]
C:Species: Eubacterium acidaminophilum
C>Date: 06-Nov-1998 #sequence_revision 06-Nov-1998 #text_change 04-Feb-2000
C:Accession: S38988; D35156
R:Luebbers, M.; Andreessen, J.R.
Eur. J. Biochem. 217, 791-798, 1993
A:Title: Components of glycine reductase from Eubacterium acidaminophilum. Cloning, s
A:Reference number: S38988; M10430; M10430; M10430; M10430; M10430; M10430; M10430
A:Accession: S38988
A:Molecule type: DNA
A:Residues: 1-315 <LUE>
A:Cross-references: GB:I04500; NID:g2708733
R:Dietrichs, D.; Meyer, M.; Schmidt, B.; Andreessen, J.R.
J. Bacteriol. 172, 2086-2095, 1990
A:Title: Purification of NADPH-dependent electron-transferring flavoproteins and N-te
A:Reference number: A35156; M10430; M10430; M10430; M10430; M10430; M10430; M10430

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:12 ; Search time 78.64 Seconds

(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-75

Perfect score: 14

Sequence: 1 SAVHDPGREGREY 14

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

PIR:68:*
1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	161	2	probable dCTP deam
2	7	50.0	1197	1	DNA-binding protei
3	6	42.9	259	2	probable exonuclea
4	6	42.9	634	2	hypothetical prote
5	6	42.9	1025	2	collagen alpha 1(V
6	5	35.7	69	2	auxin-inducible pr
7	5	35.7	88	1	immediate-early-5
8	5	35.7	119	2	50S ribosomal, prot
9	5	35.7	136	2	probable regulator
10	5	35.7	140	2	hypothetical prote
11	5	35.7	140	2	hypothetical prote
12	5	35.7	147	2	MHC class I protei
13	5	35.7	148	2	isotocin 2 / neuro
14	5	35.7	149	2	conserved hypotet
15	5	35.7	149	2	c-Jun leucine zipper
16	5	35.7	154	2	isotocin 1 / neuro
17	5	35.7	154	2	isoprenylated prot
18	5	35.7	161	2	hypothetical prote
19	5	35.7	161	2	hypothetical prote
20	5	35.7	168	2	probable dCTP deam
21	5	35.7	190	2	thiamatin homolog
22	5	35.7	191	2	probable light-rep
23	5	35.7	192	2	hypothetical prote
24	5	35.7	208	1	L-fucose-phospha
25	5	35.7	210	2	hypothetical prote
26	5	35.7	217	2	unknown protein, 8
27	5	35.7	219	2	ribose-5-phosphate
28	5	35.7	219	2	ribosephosphate is
29	5	35.7	244	2	hypothetical prote

30	5	35.7	252	1	A54677	homeotic protein 9
31	5	35.7	255	2	F83575	conserved hypotet
32	5	35.7	256	1	A42768	homeotic protein 9
33	5	35.7	271	2	D70915	hypothetical prote
34	5	35.7	272	1	UHR02	interleukin-2 rece
35	5	35.7	275	2	G70022	iron(III) diclrat
36	5	35.7	281	2	B81747	ribosomal protein
37	5	35.7	282	2	G71484	probable s2 riboso
38	5	35.7	288	2	E70580	hypothetical prote
39	5	35.7	289	2	A83656	hypothetical prote
40	5	35.7	290	2	T34620	NADH dehydrogenase
41	5	35.7	290	2	T36205	hypothetical prote
42	5	35.7	305	2	A86046	sepo (imported) -
43	5	35.7	308	2	D83452	probable cytochrom
44	5	35.7	309	2	T47685	probable RNA bindi
45	5	35.7	311	2	C84101	hypothetical prote

ALIGNMENTS

RESULT 1
E64437
probable dCTP deaminase (EC 3.5.4.13) M01102 [similarity] - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: E64437
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.I.; Overbeek, R.; Kirnness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klank, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; M01D:96337939
A:Accession: E64437
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-161 <BUL>
A:Cross-references: GB:067553; GB:L77117; NID:91591744; PIDN:MB99105.1; PID:91591747
C:Genetics:
A:Map position: FOR1043233-1043718
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SAVHDPGREGREY 14
DB 110 SAVHDPGREGREY 123

RESULT 2
A48350
DNA-binding protein - human herpesvirus 2
C:Species: human herpesvirus 2
A>Note: host Homo sapiens (man)
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 31-May-1996
C/Accession: A48350
R:Toh, Y.; Iku, Y.; Tanaka, S.; Mori, R.
Arch. Virol. 129, 183-196, 1993
A>Title: Nucleotide sequence of the major DNA-binding protein gene of herpes simplex
A:Reference number: A48350; M01D:93228441
A:Accession: A48350
A:Molecule type: DNA
A:Residues: 1-1197 <TOH>
A>Note: sequence extracted from NCBI backbone (NCBIN:129069, NCBI:129070)
C:Genetics:
A:Map position: 0.375-0.405
C:Superfamily: herpesvirus DNA-binding protein
C:Keywords: DNA binding; zinc finger

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:20 ; Search time 46.78 Seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-76
Perfect score: 14
Sequence: 1 PTIVDAGFEGLTI 14

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	173	1	DCD_ACIAM
2	11	78.6	181	1	DCD_AERPE
3	8	57.1	180	1	DCD_AOQAE
4	7	50.0	328	1	VGH_BPS13
5	7	50.0	984	1	T3RE_SALTY
6	6	42.9	258	1	ARV1_FELCA
7	6	42.9	268	1	DUT_PRYKA
8	6	42.9	287	1	ARV1_CHICK
9	6	42.9	290	1	ARV1_CHICK
10	6	42.9	290	1	ARV1_HUMAN
11	6	42.9	290	1	ARV1_HUMAN
12	6	42.9	290	1	ARV1_MESAU
13	6	42.9	290	1	ARV1_MOUSE
14	6	42.9	290	1	ARV1_MOUSE
15	6	42.9	290	1	ARV1_MOUSE
16	6	42.9	326	1	ODPB_MYGE
17	6	42.9	327	1	ODPB_MYGE
18	6	42.9	327	1	ODPB_MYCPN
19	6	42.9	328	1	VGH_BPHX
20	6	42.9	357	1	YFOL_SCHPO
21	6	42.9	440	1	YFOL_SCHPO
22	6	42.9	525	1	SYK_DEIRA
23	6	42.9	578	1	VACB_YEAST
24	6	42.9	587	1	BARI_YEAST
25	6	42.9	987	1	YMK3_YEAST
26	6	42.9	1001	1	YMK3_YEAST
27	6	42.9	1102	1	YMK3_YEAST
28	6	42.9	1114	1	YMK3_YEAST
29	6	42.9	2029	1	LAR_DROME
30	5	35.7	72	1	NIFT_AZOVI
31	5	35.7	122	1	RL7_DEIRA
32	5	35.7	130	1	RL32_THEAC
33	5	35.7	133	1	PA23_OXYSC

34	5	35.7	133	1	PA2D_PSETE
35	5	35.7	142	1	OSMC_ECOLI
36	5	35.7	143	1	14KD_MYCTU
37	5	35.7	146	1	AP4A_HUMAN
38	5	35.7	146	1	AP4A_PIG
39	5	35.7	147	1	RL15_SYNY3
40	5	35.7	149	1	GLB1_MORNR
41	5	35.7	149	1	GLB2_MORNR
42	5	35.7	149	1	GLB3_PETMA
43	5	35.7	149	1	GLB5_PETMA
44	5	35.7	149	1	GLB5_PETMA
45	5	35.7	149	1	GLB5_PETMA

ALIGNMENTS

RESULT	ID	DCD_ACIAM	STANDARD	PRT	173 AA
AC	002103				
DT	01-APR-1993 (Rel. 25, Created)				
DT	01-APR-1993 (Rel. 25, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE)				
GN	DCD				
OS	Archaeus ambivalens (Desulfurolobus ambivalens)				
OC	Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Acidilobus				
OX	NCBI_TaxID=2283				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Le1 10 / DSM 3772				
RX	MEDLINE=93065206; PubMed=1437556				
RA	Kletzin A.				
RT	"Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon Desulfurolobus ambivalens shows close phylogenetic relationship to eukaryotic ligases."				
RL	Nucleic Acids Res. 20:5389-5396(1992).				
RN	(2)				
RP	SIMILARITY				
RX	MEDLINE=95206934; PubMed=7899076				
RA	Ouzounis C., Kyriakides N., Sander C.				
RT	"Novel protein families in archaeal genomes."				
RL	Nucleic Acids Res. 23:565-570(1995).				
CC	-1- CATALYTIC ACTIVITY: DCTP + H(2)O -> DCTP + NH(3).				
CC	-1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: X63438; CAA45033.1				
DR	PIR: S26382; S26382				
DR	InterPro: IPR003232; DCTP_deaminase				
DR	InterPro: IPR001428; dctpase				
DR	Pfam: PF00692; dctpase; 1				
KW	Prodom: PD004900; dctp_deaminase; 1				
KW	Hydrolyase				
SO	SEQUENCE 173 AA; 19858 MW; B4D922503CD4B25A CRC64;				

Query Match

Best local similarity 100.0%; Score 14; DB 1; Length 173;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTIVDAGFEGLTI 14
DB 109 PTIVDAGFEGLTI 122

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:21 ; Search time 46.78 Seconds

(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-77

Sequence: 1 AHRIDPGMSGICVL 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	193	1 DCD_ECOLI	P28248 escherichia
2	8	57.1	193	1 DCD_BUCAL	P57209 buchnera ap
3	8	57.1	194	1 DCD_PASMO	P57891 pasteurella
4	8	57.1	195	1 DCD_HABIN	P44334 haemophilus
5	6	50.0	193	1 DCD_BUCAP	Q92088 buchnera ap
6	6	42.9	141	1 Y085_BORBU	O51112 varicella bu
7	6	42.9	340	1 CELE_VZDV	P09261 varicella-z
8	6	42.9	733	1 YFJ2_YEAST	P43602 saccharomyc
9	6	42.9	2444	1 MTCL_HUMAN	P46331 homo sapien
10	5	35.7	78	1 IFIC_MARPO	P12134 marchantia
11	5	35.7	107	1 FER_PSAIA	P34806 psalteriomo
12	5	35.7	132	1 YJIV_ECOLI	P39394 escherichia
13	5	35.7	141	1 HB44_XENLA	P06637 xenopus lae
14	5	35.7	142	1 HBAC_ANGAN	P80726 anguilla an
15	5	35.7	191	1 PCDB_HUMAN	O75340 homo sapien
16	5	35.7	191	1 PCDB_MOUSE	P12815 mus musculu
17	5	35.7	216	1 CSGB_SALTY	O54294 salmonella
18	5	35.7	227	1 NABC_RHOSH	O53178 rhizobium s
19	5	35.7	245	1 Y4PL_RHISN	P55617 rhizobium s
20	5	35.7	249	1 COMB_RHOSO	O53138 rhodococcus
21	5	35.7	268	1 Y064_MYCTU	O53191 rhodococci
22	5	35.7	294	1 YAFJ_HABIN	P44098 haemophilus
23	5	35.7	298	1 EFTS_MYCPN	P78009 mycoplasma
24	5	35.7	323	1 Y370_MYCGE	P47610 mycoplasma
25	5	35.7	326	1 Y370_MYCPN	P75330 mycoplasma
26	5	35.7	362	1 MLJB_HUMAN	P49286 homo sapien
27	5	35.7	410	1 B3_USTRA	P22017 ustulago ma
28	5	35.7	470	1 RHSA_RHIME	Q92312 rhizobium m
29	5	35.7	512	1 REG24_SEPLY	O13597 septoria ly
30	5	35.7	549	1 TEGU_HCMV	P07387 human cytom
31	5	35.7	663	1 RGPL_YEAST	P16664 saccharomyc
32	5	35.7	668	1 COAT_FCVF6	P27404 feline cali
33	5	35.7	668	1 COAT_FCVF4	P27405 feline cali

ALIGNMENTS

```

RESULT 1
ID DCD_ECOLI STANDARD; PRT; 193 AA.
AC P28248;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR DUS OR PAXA OR B2065.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN-K12;
RX MEDLINE-92380941; PubMed-1324907;
RA Wang L., Weiss B.;
RT "dcd (dctp deaminase) gene of Escherichia coli: mapping, cloning,
RT sequencing, and identification as a locus of suppressors of lethal
RT dut (dnpase) mutations."
RL J. Bacteriol. 174:5647-5653(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shaio Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97251358; PubMed-9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitagawa M., Kitagawa M.,
RA Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map."
RL DNA Res. 3:379-392(1996).
RN [4]
RP SUBUNIT: HOMOTETRAMER (PROBABLY).
RT DNA Res. 3:379-392(1996).
RN [5]
RP SUBUNIT: BELONGS TO THE DCTP DEAMINASE FAMILY.
RT DNA Res. 3:379-392(1996).
RN [6]
RP SUBUNIT: BELONGS TO THE DCTP DEAMINASE FAMILY.
RT DNA Res. 3:379-392(1996).

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@lsb-sib.ch).
CC EMBL: M90069; AAA23669.1;
DR EMBL: AE000296; AAC75126.1;
DR EMBL: D90844; BAA15918.1;
DR EMBL: D90845; BAA15923.1;
DR EMBL: A42940; A42940.
DR EMBL: E01118; A42940.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KM HydroLase: Complete proteome.
SQ SEQUENCE 193 AA; 21249 MW; B0044051ADE7F919 CRC64;

Query Match
Best Local Similarity 100.0%; Score 14; DB 1; Length 193;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AHRIDPGWGSCIVL 14
124 AHRIDPGWGSCIVL 137

RESULT 2
DCD_BUCAI STANDARD; PRT; 193 AA.
ID DCD_BUCAI
AC P57209;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR BU108.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Nature 407:81-86(2000).
CC -1 CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -1 SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AP001118; BAB12827.1; ALT_INIT.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KM HydroLase: Complete proteome.
SQ SEQUENCE 193 AA; 21951 MW; DE7DD7FD961FC9FA CRC64;

Query Match
Best Local Similarity 57.1%; Score 8; DB 1; Length 193;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AHRIDPGW 8
124 AHRIDPGW 131

13:32
Scott
Fitzma
Jr.

DB 124 AHRIDPGW 131
RESULT 3
DCD_PASMU STANDARD; PRT; 194 AA.
ID DCD_PASMU
AC P57891;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE).
GN DCD OR PM0951.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang O., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1 CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -1 SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AE006134; AAK03035.1;
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KM HydroLase: Complete proteome.
SQ SEQUENCE 194 AA; 21467 MW; AC76D219F5258F94 CRC64;

Query Match
Best Local Similarity 57.1%; Score 8; DB 1; Length 194;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AHRIDPGW 8
124 AHRIDPGW 131

RESULT 4
DCD_HAEIN STANDARD; PRT; 195 AA.
ID DCD_HAEIN
AC P44534;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR H1013.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / RW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Kelschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Feilavage A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:22 ; Search time 46.78 Seconds

(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-78

Sequence: 1 VGLIDSDYQGLMT 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	14	100.0	151 1 DUT_ECOLI	P06968 escherichia
2	14	100.0	154 1 DUT_BUCAL	P57623 buchnera ap
3	10	71.4	151 1 DUT_HAEN	P43792 haemophilus
4	10	71.4	151 1 DUT_PASBU	P57914 pasteurella
5	10	71.4	152 1 DUT_COXHU	P45920 coxiella bu
6	6	42.9	293 1 YMO2_RHIME	P49306 rhicobium m
7	6	42.9	586 1 HO_YEAST	P09932 saccharomyc
8	6	42.9	805 1 SYFB_MYCPN	P75563 mycoplasma
9	6	42.9	848 1 AMPN_LACIA	P48656 lactococcus
10	6	42.9	1071 1 VAIK_YEAST	P17255 saccharomyc
11	5	35.7	69 1 PAP2_SPKA	P23334 swinepox vi
12	5	35.7	98 1 YGAV_ECOLI	P77295 escherichia
13	5	35.7	128 1 YJ22_YEAST	P47093 saccharomyc
14	5	35.7	141 1 DUT_CHVPI	P41033 parametium
15	5	35.7	142 1 DUT_SPKA	P32208 swinepox vi
16	5	35.7	143 1 YL23_ARCFU	P028157 archaeoglob
17	5	35.7	145 1 DUT_CHLPPN	P09392 chlamydia p
18	5	35.7	146 1 DUT_ZYMO	P09392 chlamydia m
19	5	35.7	148 1 PBT1_DROME	P48500 bacterioph
20	5	35.7	148 1 PBT1_DROME	P54191 oryza pseu
21	5	35.7	160 1 SP26_BACTK	P26767 bacillus th
22	5	35.7	161 1 DUT_ADEG8	P09390 avian adeno
23	5	35.7	163 1 DUT_ADEG8	P05254 xenopus lae
24	5	35.7	175 1 HSLV_XENLA	P57115 buchnera ap
25	5	35.7	175 1 HSLV_XENLA	P9465 mycobacteri
26	5	35.7	188 1 RECK_MYCSM	P57123 haemophilus
27	5	35.7	199 1 YEF7_HAEIN	P97769 mus musculi
28	5	35.7	203 1 MSGL_MOUSE	P55593 synchocyst
29	5	35.7	212 1 KTHX_SYNY3	P05593 synchocyst
30	5	35.7	216 1 EGPH_HUMAN	P19748 mus musculi
31	5	35.7	216 1 EGPH_MOUSE	P19748 mus musculi
32	5	35.7	220 1 PAP2_CAPVK	P39590 bacillus su
33	5	35.7	241 1 YMBG_BACSU	

34	5	35.7	251 1 YSC1_STRGC	P42360 streptococ
35	5	35.7	261 1 TFXG_RHILT	P42729 rhizobium i
36	5	35.7	285 1 YAOV_XENLA	P09175 xenopus lae
37	5	35.7	285 1 YAOV_RHISN	P55607 rhizobium s
38	5	35.7	301 1 SCER_ZYMO	P03417 zymomonas m
39	5	35.7	307 1 DHSY_METTH	P02620 methanobact
40	5	35.7	308 1 Y117_HELPJ	P09206 helicobacte
41	5	35.7	309 1 SP26_BACSU	P56080 helicobacte
42	5	35.7	310 1 HFLC_BUCAL	P13801 bacillus su
43	5	35.7	316 1 RIAO_PLAF8	P57630 buchnera ap
44	5	35.7	317 1 FABH_RICPR	P09460 plasmodium
45	5	35.7	317 1 FABH_RICPR	P09460 plasmodium

ALIGNMENTS

RESULT 1
DUT_ECOLI STANDARD: PRT; 151 AA.
AC P06968;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYRIBIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
GN (DUTPASE) (DUTP PYROPHOSPHATASE).
DN DUT OR DNAS OR SOF OR B3640.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=84057777; PubMed=6139280;
RA Lundberg L.G., Thoreson H.-O., Karlstrom O.H., Nyman P.O.;
RT "Nucleotide sequence of the structural gene for dUTPase of
RL Escherichia coli K-12";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=93315143; PubMed=766882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RL genome: organizational symmetry around the origin of replication";
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=92158084; PubMed=1311056;
RA Cedergren-Zeppezauer E.S., Larsson G., Nyman P.O., Dauter Z.,
RT "Wilson K.S.;
RL "Crystal structure of a dUTPase";
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=96227973; PubMed=8646539;
RA Larsson G., Stenstrom L.A., Nyman P.O.;
RT "Crystal structure of the Escherichia coli dUTPase in complex with a
RL substrate analogue (dUDP)";
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=98437602; PubMed=9757098;
RA Dauter Z., Wilson K.S., Larsson G., Nyman P.O.;
RT "The refined structure of dUTPase from Escherichia coli";
RL Acta Crystallogr. D 54:735-749(1998).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.05 ANGSTROMS).
RX PubMed=11375495;
RA Gonzalez A., Larsson G., Persson R., Cedergren-Zeppezauer E.S.;
RT "Atomic resolution structure of Escherichia coli dUTPase determined ab

RT Acta Crystallogr D 57:767-774(2001).
CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUMP SO THAT
CC URACIL CANNOT BE INCORPORATED INTO DNA.
CC -1- CATALYTIC ACTIVITY: DUMP + H(2)O -> DUMP + PYROPHOSPHATE.
CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X01714; CAA25859.1; -
CC EMBL: V01578; CAA24897.1; -
CC EMBL: L10328; AAA61993.1; -
CC EMBL: AE000441; AAC76664.1; -
CC PIR: Q00497; WPECDD;
CC PDB: 1DUP; 01-SEP-95.
CC PDB: 1DDP; 08-NOV-96.
CC PDB: 1EUM; 03-MAY-00.
CC SWISS-2DPAGE: P06968; COLI.
CC ECO2DBASE: C017.2; 6TH EDITION.
CC Ecogene: EG10251; dut.
CC InterPro: IPR001428; dutPase.
CC Pfam: PF00692; dutPase; 1.
CC Hydrolase: Nucleotide metabolism: 3D-structure: Complete proteome.
CC SEQUENCE 151 AA; 16155 MW; 98FA3DE0BC70FEB2 CRC64;

Query Match 100.0%; Score 14; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGLIDSDYOGQLMI 14
DB 85 VGLIDSDYOGQLMI 98

RESULT 2
DUT_BUCAT STANDARD: PRT; 154 AA;
20-NOV-2001 (Rel. 40; Created)
20-NOV-2001 (Rel. 40; Last sequence update)
20-NOV-2001 (Rel. 40; Last annotation update)
DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
(DUTPASE) (DUTP PYROPHOSPHATASE).
DUT OR BU560.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-TOKYO 1998;
RA MEDLINE=20445173; PubMed=1093077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUMP SO THAT
CC URACIL CANNOT BE INCORPORATED INTO DNA (BY SIMILARITY).
CC CATALYTIC ACTIVITY: DUMP + H(2)O -> DUMP + PYROPHOSPHATE.
CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32776; AAC22615.1; -

CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP001119; BAB13250.1; -
CC InterPro: IPR001428; dutPase.
CC Pfam: PF00692; dutPase; 1.
CC Hydrolase: Nucleotide metabolism: Complete proteome.
CC SEQUENCE 154 AA; 16936 MW; E5B5FDEBF809E920 CRC64;

Query Match 100.0%; Score 14; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VGLIDSDYOGQLMI 14
DB 85 VGLIDSDYOGQLMI 98

RESULT 3
DUT_HAEIN STANDARD: PRT; 151 AA;
01-NOV-1995 (Rel. 32; Created)
01-NOV-1995 (Rel. 32; Last sequence update)
20-NOV-2001 (Rel. 40; Last annotation update)
DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
(DUTPASE) (DUTP PYROPHOSPHATASE).
DUT OR HI0954.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-RD / KW20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fiedlschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bait C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodex A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUMP SO THAT
CC URACIL CANNOT BE INCORPORATED INTO DNA (BY SIMILARITY).
CC CATALYTIC ACTIVITY: DUMP + H(2)O -> DUMP + PYROPHOSPHATE.
CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32776; AAC22615.1; -

DR HSP: P06968
InterPro: H10
Pfam: 1
EC: 3.6.1.23

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:23 ; Search time 46.78 Seconds

(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-79

Perfect score: 14
Sequence: 1 AGVDRDYTGKVK 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	147	1 DUT_YEAST	P33317 saccharomyc
2	6	42.9	123	1 ATPE_HELPJ	Q92K82 helicobacte
3	6	42.9	123	1 ATPE_HELPJ	P56084 helicobacte
4	6	42.9	132	1 MERR_BACSR	P22853 bacillus sp
5	6	42.9	185	1 VGI6_HABIN	P71387 haemophilus
6	6	42.9	316	1 TALB_ECOLI	P30148 escherichia
7	6	42.9	491	1 TRPE_NETGO	Q9W400 neisseria g
8	6	42.9	491	1 TRPE_NETMA	Q9X420 neisseria m
9	6	42.9	491	1 TRPE_NETMB	P56955 neisseria m
10	6	42.9	491	1 TRPE_NETMC	Q9S358 neisseria m
11	6	42.9	1434	1 RPOB_UREPA	Q9P456 ureaplasma
12	5	35.7	67	1 Y737_ARCFU	Q29521 archaeoglob
13	5	35.7	72	1 VGI8_BPMU	P25955 bacterioph
14	5	35.7	98	1 CMGC_BACSU	P25955 bacterioph
15	5	35.7	110	1 CUI6_MANSE	Q25504 manduca sex
16	5	35.7	111	1 KVI3_MOUSE	P01604 mus musculu
17	5	35.7	113	1 PURL_SYNPF	Q50413 synecococc
18	5	35.7	114	1 DCHS_LACBU	P04193 lactobacill
19	5	35.7	124	1 RSI2_TREPA	Q83271 treponema p
20	5	35.7	128	1 LRP_ORYSA	Q03200 oryza sativ
21	5	35.7	129	1 ATP_VIBAL	P12963 vibrio algi
22	5	35.7	141	1 AMPM_MYCA	Q59509 mycoplasma
23	5	35.7	143	1 LGB1_VICFA	P02232 vicia faba
24	5	35.7	149	1 RL19_ARCFU	Q28372 archaeoglob
25	5	35.7	150	1 DUT_AQUAE	Q66592 aquifex aeo
26	5	35.7	158	1 SODC_ONCVO	P24706 onchocerca
27	5	35.7	159	1 DUT_CANAL	P43068 candida alb
28	5	35.7	162	1 Y269_AQUAE	Q66626 aquifex aeo
29	5	35.7	167	1 NUGM_MACRO	P92670 macropteryc
30	5	35.7	168	1 YNP7_YEAST	P53897 saccharomyc
31	5	35.7	175	1 RIMM_PSEAE	Q9HX40 pseudomonas
32	5	35.7	179	1 YXAE_CYPAA	P48335 cyanophora
33	5	35.7	181	1 NDKM_COLL1	P87355 columba liv

34	5	35.7	186	1 CHS2_USTMA	P30599 ustillago ma
35	5	35.7	192	1 EPO_MOUSE	P07321 mus musculu
36	5	35.7	193	1 H1S7_SULSO	Q33773 sulfolobus
37	5	35.7	202	1 VMAF_RABVN	P25224 rabies viru
38	5	35.7	214	1 VCOL_VARY	P33859 variola vir
39	5	35.7	221	1 FIXW_RHILE	P14312 rhizobium l
40	5	35.7	226	1 TPIS_RHLET	P96985 rhizobium e
41	5	35.7	228	1 C79B_MOUSE	P15530 mus musculu
42	5	35.7	239	1 C79B_HUMAN	P40259 homo sapien
43	5	35.7	234	1 Y352_JHEMA	Q9W417 thermotoga
44	5	35.7	238	1 Y206_CHLMD	Q9P141 chlamydia m
45	5	35.7	239	1 RNL2_LENED	P81296 lentinula e

ALIGNMENTS

RESULT 1
DUT_YEAST STANDARD; PRT; 147 AA.
AC P33317;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPASE) (DUTP PYROPHOSPHATASE).
GN DUT1 OR YBR252W OR YBR1705.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94038925; PubMed=8223452;
RA Gadsden M.H., McIntosh E.M., Game J.C., Wilson P.J., Haynes R.H.;
RT "DUTP pyrophosphatase is an essential enzyme in Saccharomycetes
cerevisiae.";
RL EMBL J. 12:4425-4431(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94078675; PubMed=8256522;
RA Dolignon F., Bileau N., Aigle M., Grouzet M.;
RT "The complete sequence of a 6794 bp segment located on the right arm
of chromosome II of Saccharomycetes cerevisiae. Finding of a putative
dutpase in a yeast.";
RL Yeast 9:1131-1137(1993).
CC - FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
URACIL CANNOT BE INCORPORATED INTO DNA.
CC - CATALYTIC ACTIVITY: DUTP + H(2)O -> DUMP + PYROPHOSPHATE.
CC - PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
CC - SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL: X74263; CAA52322.1; -
DR EMBL: L20286; AAA5611.1; -
DR EMBL: Z36121; CAA85215.1; -
DR PIR: S38189; S38189.
DR SGD: S0000456; DUT1.
DR InterPro: IPR001428; dutpase.
DR Pfam: PF00692; dutpase; 1.
KW Hydroxylase; Nucleotide metabolism.
FT CONFLICT 10 N -> K (IN REF. 2).
FT SEQUENCE 147 AA; 15293 MW; 6F1E87A692A061F6 CRC64;

Query Match 100.0%; Score 14; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 6e-08; 0;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVYDRDYTEGVK 14
 |||||
 DB 81 AGVYDRDYTEGVK 94

RESULT 2

ATPE_HELPJ STANDARD; PRT; 123 AA.

ID ATPE_HELPJ
 AC 092K82;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).
 DE ATPC OR JHP1059.
 Helicobacter pylori J99 (Campylobacter pylori J99).
 Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 Helicobacter.
 NCBI_Taxid=85963;
 OX [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deGange B.L., Carmel G.,
 RA Tummino P.J., Carnuso A., Ulla-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Weidberg D., Mills S.D., Jiang Q., Taylor D.E., Vois G.F.,
 RA Trust J.J.,
 RT Genomic sequence comparison of two unrelated isolates of the human
 gastric pathogen Helicobacter pylori.
 RL Nature 357:176-180(1999).
 CC - FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 GRADIENT ACROSS THE MEMBRANE.
 CC - SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 HAS THREE MAIN SUBUNITS: A, B AND C.
 CC - SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL: AEO01533; AAD06638.1;
 DR HSSP: P00832; 1BSH
 DR InterPro: IPR001469; ATP-synt_DE.
 DR Pfam: PF00401; ATP-synt_DE; 1.
 DR ProDom: PD000944; ATP-synt_DE; 1.
 DR HydroLase: ATP synthesis; CF(1); Hydrogen ion transport;
 KW Complete proteome.
 SO SEQUENCE 123 AA; 13271 MW; 19F561ED65B49761 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 123;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YTGGEVK 13
 |||||
 DB 16 YTGGEVK 21

STANDARD; PRT; 123 AA.

NCBI Taxid=85963;
 OX [1]
 RN Bacillus/sta
 MEDLINE
 WU

AC P56084;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).
 GN ATPC OR HPI131.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 NCBI_Taxid=210;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=26695 / ATCC 700392;
 RX MEDLINE=9739467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.,
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori."
 RL Nature 388:539-547(1997).
 CC - FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 GRADIENT ACROSS THE MEMBRANE.
 CC - SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 HAS THREE MAIN SUBUNITS: A, B AND C.
 CC - SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL: AEO00619; AAD08173.1; ALT_INT.
 DR HSSP: P00832; 1AOT.
 DR TRIG: HP1131;
 DR InterPro: IPR001469; ATP-synt_DE.
 DR Pfam: PF00401; ATP-synt_DE; 1.
 DR ProDom: PD000944; ATP-synt_DE; 1.
 DR HydroLase: ATP synthesis; CF(1); Hydrogen ion transport;
 KW Complete proteome.
 SO SEQUENCE 123 AA; 13357 MW; 19EDDAD2B3A5461 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 123;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YTGGEVK 13
 |||||
 DB 16 YTGGEVK 21

RESULT 4
 MERR_BACSR STANDARD; PRT; 132 AA.

AC P22853;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE MERCURIC RESISTANCE OPERON REGULATORY PROTEIN.
 GN MERR.
 OS Bacillus sp. (strain RC607).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:24 ; Search time 46.78 seconds

(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-80

Sequence: 1 AGVIDEDYRGNGV 14

Scoring table:

GAPD 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	14	100.0	159	1 DUT_ORFN2
2	14	100.0	205	1 DUT_RAT
3	14	100.0	252	1 DUT_HUMAN
4	10	71.4	144	1 DUT_VACCV
5	10	71.4	147	1 DUT_VACCC
6	10	71.4	147	1 DUT_VARY
7	7	50.0	1003	1 MYSE_DICDI
8	7	50.0	2021	1 OMPA_RICCN
9	6	42.9	171	1 PSAN_ARATH
10	6	42.9	178	1 DUT_ADEGI
11	6	42.9	218	1 IM23_SCHMA
12	6	42.9	271	1 DABP_MYCBO
13	6	42.9	289	1 YHDE_BACSU
14	6	42.9	624	1 SERA_ARATH
15	6	42.9	757	1 RRP1_IAROM
16	6	42.9	766	1 ERA_DROME
17	6	42.9	878	1 SYA_BOCAL
18	6	42.9	1365	1 YAR1_SCHPO
19	6	42.9	1391	1 N157_YEAST
20	5	35.7	76	1 VA19_VARY
21	5	35.7	82	1 COAB_BRPFI
22	5	35.7	93	1 RLA0_PIG
23	5	35.7	100	1 CHA3_BOMBO
24	5	35.7	103	1 VP16_NFVOP
25	5	35.7	109	1 YS88_CAEEL
26	5	35.7	114	1 UNR_GAVPO
27	5	35.7	120	1 YHRY_BCOLI
28	5	35.7	129	1 CHA1_BOMBO
29	5	35.7	131	1 YD2E_SCHPO
30	5	35.7	132	1 CHA2_BOMBO
31	5	35.7	140	1 DUT_SCHPO
32	5	35.7	141	1 DUT_CHYPI
33	5	35.7	143	1 RK2_SOYBN

34	5	35.7	148	1 DUT_BPT5	048500 bacterioph
35	5	35.7	148	1 YFB2_YEAST	P43580 saccharomyc
36	5	35.7	148	1 YH62_YEAST	P36839 saccharomyc
37	5	35.7	157	1 UL51_HCVVA	P16792 human cytom
38	5	35.7	158	1 CRGF_MOUSE	003740 mus musculu
39	5	35.7	159	1 DUT_CANAL	P43058 candida alb
40	5	35.7	163	1 DUT_ADEG8	09Y50 avian adeno
41	5	35.7	164	1 FLAV_HELPJ	092K53 helicobacte
42	5	35.7	164	1 FLAV_HELPJ	025776 helicobacte
43	5	35.7	169	1 DUT_LYCES	P32518 lycopersico
44	5	35.7	173	1 CRGD_BOVIN	P08209 bos taurus
45	5	35.7	173	1 CRGD_HUMAN	P07320 homo sapien

ALIGNMENTS

RESULT 1	DUT_ORFN2	STANDARD;	PRT; 159 AA.
ID	DUT_ORFN2		
AC	P14597;		
DR	01-APR-1990 (Rel. 14, Created)		
DT	01-APR-1990 (Rel. 14, Last sequence update)		
DE	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	DEOXYRIBIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)		
DE	(DUTPASE) (DUTP PYROPHOSPHATASE).		
OS	Orf virus (strain NZ2) (OV NZ-2).		
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;		
OC	Parapoxvirus.		
OX	NCBI_Taxid=10259;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90021200; PubMed=2678731;		
RA	Merger A.A., Fraser K.M., Stockwell P.A., Robinson A.J.;		
RT	"A homologue of retroviral pseudoproteases in the parapoxvirus, orf		
RT	virus.";		
RL	Virology 172:665-668(1989).		
CC	-1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT		
CC	PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES		
CC	AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT		
CC	URACIL CANNOT BE INCORPORATED INTO DNA.		
CC	-1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.		
CC	-1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.		
CC	-1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PROTEASE-LIKE PROTEIN		
CC	(PSEUDOPROTEASE).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@sib-sib.ch).		
CC	-----		
DR	EMBL; M30023; AAA6786.1; -		
DR	HSSP; P06968; IDDD.		
DR	InterPro; IPR001428; DUTPase.		
DR	Pfam; PF00692; DUTPase; 1.		
KW	Hydrolase; Nucleotide metabolism.		
SQ	SEQUENCE 159 AA; 16893 MW; 08CD852D47AE17AD CRC64;		
Query Match	100.0%; Score 14; DB 1; Length 159;		
Best Local Similarity	100.0%; Pred. No. 2.9e-08;		
Matches 14; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 AGVIDEDYRGNGV 14		
DB	81 AGVIDEDYRGNGV 94		
RESULT 2	DUT_RAT		

ID DUT_RAT STANDARD: PRT: 205 AA.
 AC P70583;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
 DE (DUTPASE) (DUTP PYROPHOSPHATASE) (PPAR-INTERACTING PROTEIN 4) (PIPA)
 GN DUT.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=97066956; PubMed=8910358;
 RA Chu R.Y., Lin Y.L., Rao M.S., Reddy J.K.;
 RT "Cloning and identification of rat deoxyuridine triphosphatase as an
 RT inhibitor of peroxisome proliferator-activated receptor alpha."
 RT J. Biol. Chem. 271:27670-27676(1996).
 (2)
 REVISIONS.
 RA Chu R.Y., Lin Y.L., Rao M.S., Reddy J.K.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 CC URACIL CAN BE INCORPORATED INTO DNA. INHIBITS PEROXISOME
 CC PROLIFERATION BY PREVENTING THE LATTER'S DIMERIZATION WITH
 CC RETINOLACTIC RECEPTOR.
 CC ENZYME TO PPAR. PREVENTING THE LATTER'S DIMERIZATION WITH
 CC RETINOLACTIC RECEPTOR.
 CC CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC COFACTOR: MAGNESIUM (BY SIMILARITY).
 CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. BINDING TO PPAR INDUCES
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED. HIGHER
 CC LEVELS IN HEART AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U64030; AAC34734.2;
 CC HSSP: P16088; IDUT.
 DR InterPro: IPR001428; DUTPase.
 KM Pfam: PF00692; dUTPase; 1.
 DR Hydrolase; Nucleotide metabolism; Magnesium.
 SQ SEQUENCE 205 AA; 22003 MW; A9D54BE5ED015C4 CRC64;
 Query Match 100.0%; Score 14; DB 1; Length 205;
 Best Local Similarity 100.0%; Pred. No. 3.6e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGVIDEDYGRNGV 14
 DB 139 AGVIDEDYGRNGV 152
 RESULT 3
 ID DUT_HUMAN STANDARD: PRT: 252 AA.
 AC P33316; Q16860; Q16708; Q14785;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE PRECURSOR

DE (EC 3.6.1.23) (DUTPASE) (DUTP PYROPHOSPHATASE).
 GN DUT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (DUT-M AND DUT-N).
 RA Pearlman R.E.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (DUT-M).
 RA Lader R.D., Caradonna S.J.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (DUT-N).
 RA Cohen D., Heng H.H.O., Shi X.M., McIntosh E.M., Tsui L.C.,
 RA Pearlman R.E.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (DUT-N).
 RC TISSUE=T-cell;
 RX MEDLINE=96205967; PubMed=8631816;
 RA Lader R.D., McNulty D.E., Carr S.A., Roberts G.D., Caradonna S.J.;
 RT "Characterization of distinct nuclear and mitochondrial forms of
 RT human deoxyuridine triphosphatase nucleotidohydrolase."
 RT J. Biol. Chem. 271:7745-7751(1996).
 (5)
 RP SEQUENCE OF 112-252 FROM N.A.
 RX MEDLINE=92390380; PubMed=1325640;
 RA McIntosh E.M., Ager D.D., Gadsden M.H., Haynes R.H.;
 RT "Human dUTP pyrophosphatase: cDNA sequence and potential biological
 RT importance of the enzyme."
 RT Proc. Natl. Acad. Sci. U.S.A. 89:8020-8024(1992).
 (6)
 RP ERRATUM.
 RA McIntosh E.M., Ager D.D., Gadsden M.H., Haynes R.H.;
 RL Proc. Natl. Acad. Sci. U.S.A. 90:4328-4328(1993).
 (7)
 RP SEQUENCE OF 112-252 FROM N.A., PARTIAL SEQUENCE, AND PHOSPHORYLATION.
 RC TISSUE=Lymphocytes;
 RX MEDLINE=93281681; PubMed=8389461;
 RA Strahler J.R., Zhu X.-X., Wang Y.K., Hora N., Andrews P.C.,
 RA Roseman N.A., Neel J.V., Turka L., Hanash S.M.;
 RT "Maturation stage and proliferation-dependent expression of dUTPase
 RT in human T cells."
 RT Proc. Natl. Acad. Sci. U.S.A. 90:4991-4995(1993).
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 CC URACIL CAN BE INCORPORATED INTO DNA.
 CC CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC COFACTOR: MAGNESIUM. PHOSPHORYLATION IS NECESSARY FOR ACTIVITY.
 CC -1- ENZYME REGULATION: PHOSPHORYLATION IS NECESSARY FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: DUT-N IS NUCLEAR. DUT-M IS MITOCHONDRIAL.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: DUT-M (SHOWN HERE) AND DUT-N;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. DUT-N IS THE MORE ABUNDANTLY
 CC EXPRESSED FORM.
 CC -1- TISSUE SPECIFICITY: FOUND IN A VARIETY OF TISSUES.
 CC -1- PTM: PHOSPHORYLATION IN MATURE T-CELLS OCCUR IN A CELL CYCLE-
 CC DEPENDENT MANNER.
 CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AF016432; AB071393.1;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:25 ; Search time 46.78 Seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-81

Perfect score: 14
Sequence: 1 TGLDPERGSEKLT 14

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Swissprot_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	287	1 DUT_HSVSA	001034 herpesvirus
2	42.9	177	1	RNA5_ASPGI	P00655 aspergillus
3	42.9	186	1	NUDE_ECOLI	P45799 escherichia
4	42.9	250	1	Y122_YEAST	007953 saccharomyc
5	42.9	294	1	Y237_MYCIN	P75455 mycoplasma
6	42.9	299	1	CORC_HAETN	057368 haemophilus
7	42.9	300	1	CORC_PASMO	09cm13 pasteurella
8	42.9	308	1	PEX7_SCHPO	P78798 schizosacch
9	42.9	511	1	LNT_PSEAE	094186 pseudomonas
10	42.9	635	1	CA28_HUMAN	P02462 homo sapien
11	42.9	1214	1	STP1_HUMAN	060315 homo sapien
12	42.9	1669	1	CA14_HUMAN	P02462 homo sapien
13	42.9	1670	1	CA14_MOUSE	P02462 mus musculu
14	42.9	1670	1	CA24_MOUSE	P08122 mus musculu
15	42.9	1707	1	RODZ_CYACA	019907 cyamidium c
16	35.7	72	1	RRKZ_CYACA	P03344 rhizobium m
17	35.7	90	1	DBH_RHIME	030092 archaeoglob
18	35.7	112	1	Y145_ARCFO	P44880 haemophilus
19	35.7	143	1	USPA_ECOLI	P28242 escherichia
20	35.7	143	1	YKRS_ECOLI	P48855 bacillus su
21	35.7	172	1	APRD_BACSU	09cm27 mycobacteri
22	35.7	183	1	APRD_CYACA	007247 mycobacteri
23	35.7	190	1	DCD_MYCTU	09x8w0 streptomyce
24	35.7	191	1	DCD_STRCO	P29113 agrobacteri
25	35.7	191	1	PIC2_AGRTO	027875 methanobact
26	35.7	197	1	DCD_METHH	058298 methanococc
27	35.7	206	1	Y888_METUA	009676 saccharosac
28	35.7	219	1	Y403_SCHPO	051082 borrelia bu
29	35.7	223	1	UNG_BORBU	P57346 buchnera ap
30	35.7	226	1	RNC_BOCAT	P24247 alcaligenes
31	35.7	228	1	TFDT_ALCEU	P39590 bacillus su
32	35.7	241	1	YMBG_BACSU	P54451 bacillus su
33	35.7	243	1	YGEF_BACSU	

34	5	35.7	246	1	NCAP_SFVS	P12793 sandfly fev
35	5	35.7	250	1	YC97_HUMAN	O9y3as homo sapien
36	5	35.7	253	1	PCRB_THERC	O9h1h3 thermoplasm
37	5	35.7	254	1	FAEL_ECOLI	P33783 escherichia
38	5	35.7	254	1	GTXC_ORYSA	O06398 oryza sativ
39	5	35.7	255	1	UPPS_METHH	O26334 methanobact
40	5	35.7	263	1	KKR9_STRRI	P13250 streptomyce
41	5	35.7	263	1	YKRP_ECOLI	P75981 escherichia
42	5	35.7	264	1	DP3A_SACER	P34699 saccharopol
43	5	35.7	272	1	PSBO_SYNER	P55221 synecococc
44	5	35.7	273	1	NUFM_NEUCR	P24919 neutrospora
45	5	35.7	282	1	AQQA_ATRCA	P42767 atirplex ca

ALIGNMENTS

RESULT 1
DUT_HSVSA STANDARD; PRT; 287 AA.
ID DUT_HSVSA
AC 001034;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPase) (DUTP PYROPHOSPHATASE).
GN 54 OR EDRF3.
OS Herpesvirus salmuri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.C., Nicholson J., Bille D., Cameron K.R., Blesinger B.,
RA Newman C., Wiltman S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus salmuri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230228; PubMed=1314457;
RA Nicholson J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
herpesvirus salmuri (HVS) L-DNA: general conservation of genetic
organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
CC -I- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM; IT
PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
URACIL CANNOT BE INCORPORATED INTO DNA.
CC -I- CATALYTIC ACTIVITY: DUTP + H(2)O -> DUMP + PYROPHOSPHATE.
CC -I- SIMILARITY: BELONGS TO THE DUTPase FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.jsb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X64346; CAA4567.1; -;
DR EMBL: M86409; AAA46131.1; -;
DR PIR: G36811; WZBPPL
DR InterPro: IPR001428; DUTPase.
DR Pfam: PF00692; DUTPase; 1.
KW Hydrolyase; Nucleotide metabolism.
SEQUENCE 287 AA; 32507 MW; 598D495D74274A11 CRC64;

Query Match 100.0%; Score 14; DB 1; Length 287;
Best Local Similarity 100.0%; Pred. No. 1; Se-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGLIDPGFQGLKLT 14
 DB 74 TGLIDPGFQGLKLT 87

RESULT 2
 RNAS_ASPGI STANDARD: PRT: 177 AA.
 AC P00655;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE RIBONUCLEASE ALPHA-SARCIN PRECURSOR (EC 3.1.27.10).
 GN SAR.
 OS Aspergillus giganteus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
 NCBI_TaxID=5060;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=90245591; PubMed=2336369;
 Oka T., Natori Y., Tanaka S., Tsurugi K., Endo Y.;
 "Complete nucleotide sequence of cDNA for the cytotoxin alpha
 sarcin.";
 RT Nucleic Acids Res. 18:1897-1897(1990).
 RN [2]
 SEQUENCE FROM N.A.
 RC STRAIN-MDH 1894;
 RA Wundt S., Felske H., Henze P.F., Ulbrich N., Stahl U.;
 Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 SEQUENCE OF 28-177
 MEDLINE=83213554; PubMed=6343394;
 Sacco G., Dickmer K., Wool T.G.;
 "The primary structure of the cytotoxin alpha-sarcin.";
 RT J. Biochem. 258:5811-5818(1983).
 CC -1- FUNCTION: ALPHA-SARCIN IS SPECIFIC FOR PURINES IN BOTH SINGLE- AND
 J. FUNCTION: ALPHA-SARCIN IS SPECIFIC FOR PURINES IN BOTH SINGLE- AND
 CC -1- FUNCTION: STRANDED RNA. ITS TOXIC ACTION ON EUKARYOTIC CELLS IS THE
 CC RESULT OF CLEAVAGE OF A SINGLE PHOSPHODIESTER BOND IN THE 60S
 CC SUBUNIT OF RIBOSOMES.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE PHOSPHODIESTER LINKAGE
 CC BETWEEN GUANOSINE AND ADENOSINE RESIDUES AT ONE SPECIFIC POSITION
 CC IN THE 28S RRNA FROM RAT RIBOSOMES.
 CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE U2 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D13704; BAA02863.1;
 CC EMBL: X60770; CAA03180.1;
 CC DR PIR: S12582; NRASG.
 CC DR PIR: S21866; S21866.
 CC DR HSP: P04389; IAOZ.
 CC KW Hydrolyase; Nuclease; Protein synthesis inhibitor; Signal.
 CC FT SIGNAL 1 27
 CC FT CHAIN 28 177
 CC FT DISULFID 33 175
 CC FT DISULFID 103 159
 CC FT ACT_SITE 123 123
 CC FT ACT_SITE 123 123
 CC SEQUENCE 177 AA; 19724 MW; 6C711B9482DC9D1 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 177;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QGLIDPG 14
 DB 169 QGLIDPG 174

RESULT 3
 NUDE_FCOLI STANDARD: PRT: 186 AA.
 AC P45759;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADP COMPOUNDS HYDROLASE NUDE (EC 3.6.1.7).
 GN NUDE OR B3397.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 NCBI_TaxID=562;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA Blatter F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RN [2]
 CHARACTERIZATION.
 RP MEDLINE=98123081; PubMed=9452430;
 RA O'Handley S.F., Erick D.N., Dunn C.A., Bessman M.J.;
 "Orf186 represents a new member of the Nudix hydrolase, active on
 adenosine(5')triphosphate(5')adenosine, ADP-ribose, and NADH.";
 RT J. Biol. Chem. 273:3192-3197(1998).
 CC -1- FUNCTION: ACTIVE ON ADENOSINE(5')TRIPHOSPHATE(5')ADENOSINE (AP3A),
 CC ADP-RIBOSE, NADH, ADENOSINE(5')DIPHOSPHATE(5')ADENOSINE (AP2A).
 CC -1- CATALYTIC ACTIVITY: ADP-RIBOSE + H(2)O -> AMP + D-RIBOSE 5'-
 CC PHOSPHATE.
 CC -1- COFACTOR: MAGNESIUM OR OTHER DIVALENT CATIONS.
 CC -1- SUBUNIT: HOMODIMER (PROBABLY).
 CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U18997; AA58194.1;
 CC EMBL: AF000415; AAC76422.1;
 CC DR EMBL: BC12926; NUDE.
 CC DR EMBL: F000086; NUDIX_hydrolase.
 CC DR PIR: P000293; mutt; 1.
 CC DR PIR: P000502; NUDIXDOMAIN.
 CC DR PRINTS: PS00893; NUDIX; 1.
 CC DR PROSITE: PS00893; NUDIX; 1.
 CC KW Hydrolyase; Manganese; Complete proteome.
 CC FT DOMAIN 80 101
 CC FT DOMAIN 101 101
 CC SEQUENCE 186 AA; 21153 MW; D959AD8ECF73FC4 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 186;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GLIDPG 7
 DB 80 GLIDPG 85

RESULT 4

C:Keywords: hydrolase

Query Match 14.1%; Score 22; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 3,4e-15;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GSFANVDPGMDGNLTMLYNAS 110
DB 89 GSFANVDPGMDGNLTMLYNAS 110

RESULT 3

DB4965
C:Species: Buchnera sp.
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: DB4965
R:Shigenobu, S.; Matanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nucleotide: 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A:Reference number: A84930; MUID:20445173
A:Accession: DB4965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: suchB, BU303
A:Superfamily: dihydrolipoamide acetyltransferase: lipoyl/biotin-binding homology
C:Keywords: acyltransferase; coenzyme A

Query Match 5.1%; Score 8; DB 2; Length 420;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 EREKGVYI 52
DB 62 EREKGVYI 69

RESULT 4

I40398
C:Species: Bacillus subtilis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: I40398; E69625
R:Chen, L.; Helmann, J.D.
J. Bacteriol. 176, 3093-3101, 1994
A:Title: The Bacillus subtilis sigma D-dependent operon encoding the flagellar proteins
A:Reference number: I40396; MUID:94252974
A:Accession: I40398
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-133 <RES>
A:Cross-references: EMBL:Z31376; NID:g499379; PIDN:CA83249.1; PID:g499382
R:Kunst, S.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrati, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gallet
Jech, J.; Harwood, C.R.; Hensut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
Koeiler, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Riegler, M.; Rivoira, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
R.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
A:Authors: Schleich, S.; Schaefer, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot
Akeuchi, M.; Tanaka, S.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Whitters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033

A:Accession: E69625
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-133 <KUN>
A:Cross-references: GB:299122; GB:AL009126; NID:g2636029; PIDN:CA815550.1; PID:g26360
A:Experimental source: strain 168
C:Genetics:
A:Gene: flis
A:Superfamily: flagellar protein flis

Query Match 4.5%; Score 7; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LRLMLYN 108
DB 23 LRLMLYN 29

RESULT 5

EB4101
C:Species: Bacillus halodurans (strain C-125)
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000
C:Accession: EB4101
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: EB4101
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-133 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807332.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: flis
C:Superfamily: flagellar protein flis

Query Match 4.5%; Score 7; DB 2; Length 133;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LRLMLYN 108
DB 23 LRLMLYN 29

RESULT 6

A35123
all protein precursor - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C>Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 02-Sep-2000
C:Accession: A35123
R:Miller, V.L.; Biliska, J.B.; Falkow, S.
J. Bacteriol. 172, 1062-1069, 1990
A:Title: Nucleotide sequence of the Yersinia enterocolitica AII gene and characteriza
A:Reference number: A35123; MUID:90130261
A:Accession: A35123
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <MII>
A:Cross-references: GB:M29945; NID:g155437; PIDN:AA88694.1; PID:g155438
C:Superfamily: phage lambda membrane protein Iom

Query Match 4.5%; Score 7; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 GYIGSFA 92
|||||
DB 65 GYIGSFA 71

RESULT 7

alkyl hydroperoxide reductase homolog - Sulfolobus solfataricus
S74033
N/Alternate names: protein C0215
C/Species: Sulfolobus solfataricus
C/Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000
C/Accession: S74033
R/Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.
Mol. Microbiol. 22, 175-191, 1996
A/Title: Organizational characteristics and information content of an archaeal genome: I
A/Reference number: S73076; MUID:97055432
A/Accession: S74033
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-215 <SEN>
A/Cross-references: EMBL:Y08256; NID:g1707679; PIDN:CAA69447.1; PID:g1707727
A/Experimental source: strain P2
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996
C/Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase C22 prot

Query Match 4.5%; Score 7; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 RIKLPPD 71
|||||
DB 24 RIKLPPD 30

RESULT 8

C64491
phosphoribosylformimino-5-aminimidazole carboxamide ribotide isomerase - Methanococcus
C/Species: Methanococcus jannaschii
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: C64491
R/Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sedow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A/Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A/Reference number: A64300; MUID:96337999
A/Accession: C64491
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-237 <BLU>
A/Cross-references: GB:067594; GB:L77117; NID:g1592160; PIDN:AAB99553.1; PID:g1592163; T
C/Genetics:
A/Map position: FOR1508884-1509597
C/Superfamily: N-(5'-phospho-D-riboseylformimino)-5-amino-1-(5'-phosphoribosyl)-4-imidaz

Query Match 4.5%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 KEGKVI 52
|||||
DB 131 KEGKVI 137

RESULT 9

A69000
conserved hypothetical protein MTH1 - Methanobacterium thermoautotrophicum (strain Delta
C/Species: Methanobacterium thermoautotrophicum
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C/Accession: A69000

R/Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Olu, D.; Spadafora, R.; Yalcire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
K.L.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Neelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A/Reference number: A69000; MUID:98037514
A/Accession: A69000
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-268 <MTH>
A/Cross-references: GB:AE000795; GB:AE000666; NID:g2621036; PIDN:AAB84510.1; PID:g262
A/Experimental source: strain Delta H
C/Genetics:
A/Gene: MTH1
C/Superfamily: conserved hypothetical protein MTH1

Query Match 4.5%; Score 7; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 KEILIEP 16
|||||
DB 149 KEILIEP 155

RESULT 10

A64149
hypothetical protein HI0345 - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C/Accession: A64149
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A/Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MUID:95350630
A/Accession: A64149
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-279 <TIGR>
A/Cross-references: GB:U32719; GB:L42023; NID:g1573310; PIDN:ABC22006.1; PID:g1573315
A/Note: best homolog was a hypothetical protein from Escherichia coli
C/Superfamily: conserved hypothetical protein HI0345; ferredoxin 2[4Fe-4S] homology
F;65-130/Domain: ferredoxin 2[4Fe-4S] homology <FER>

Query Match 4.5%; Score 7; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 SLAREGV 87
|||||
DB 48 SLAREGV 54

RESULT 11

A70435
cysteine synthase (EC 4.2.99.8) cysM (similarity) - Aquifex aeolicus
C/Species: Aquifex aeolicus
C/Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Jun-2000
C/Accession: A70435
R/Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A/Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A/Reference number: A70300; MUID:98196666
A/Accession: A70435
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-327 <AQF>

DE 01-NOV-1999 (TREMblrel. 12, last annotation update)
 DE K08C9.6 PROTEIN.
 GN Caenorhabditis elegans.
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 CC NCB1_TaxID=6239;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP Lennard N.;
 RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Atsough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Cration M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifkin C., Roopra A., Saunders D., Shownkeen R.,
 Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 Thelery-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 Watson A., Wellstock L., Wilkinson-Sproat J., Woldman P., III of C.
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans";
 RL Nature 368:32-38(1994).
 DR EMBL: Z81567; CAB04589.1; -;
 SQ SEQUENCE 127 AA; 14638 MW; B169A1E75006C2CE CRC64;

Query Match 2.0%; Score 8; DB 5; Length 127;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 248 FRIRIKL 255
 DB 85 FRIRIKL 92

RESULT 10
 09RLAT PRELIMINARY; PRT; 142 AA.
 AC 09RLAT; (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE SODIUM-HYDROGEN EXCHANGER REGULATORY FACTOR (FRAGMENT).
 GN SLC9A3RL.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SV;
 RA Weinman E.J., Stepiak D.A., Zhang X., Akhter S., Shenolikar S.;
 RT "Molecular cloning of the CDNA and promoter sequences for the mouse
 sodium-hydrogen exchanger regulatory factor.";
 RL Biochim. Biophys. Acta 0:0-0(1999).
 DR EMBL: AF154912; AAD49224.1; -;
 DR HSP: Q12923; 3PDZ.
 DR MGD: MGI:1349482; SLC9A3RL.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PSS0106; PDZ; 1.
 FT NON-TER 142
 SQ SEQUENCE 142 AA; 15376 MW; A8994D6A865B283A CRC64;

Match Similarity 2.0%; Score 8; DB 11; Length 142;
 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 152 ERKRLGV 159
 DB 97 ERKRLGV 104

RESULT 11
 09ZSW7 PRELIMINARY; PRT; 169 AA.
 ID 09ZSW7
 AC 09ZSW7;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE BETA-FRUCTOFURANOSIDASE (EC 3.2.1.26) (FRAGMENT).
 OS Hamamelis virginiana.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Saxifragales; Hamamelidaceae; Hamamelis.
 OX NCB1_TaxID=4397;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LEAF;
 RA Szpara M.L.T.;
 RT "Isolation of invertase gene sequences from witch hazel (Hamamelis
 virginiana).";
 RL Thesis (1998); Pennsylvania State University.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING BETA-D-
 FRUCTOFURANOSIDE RESIDUES IN BETA-D-FRUCTOFURANOSIDES.
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
 CC EMBL: AF091547; AAC99430.1; -;
 DR Mendel: 36382; Hamyl: 1002; 36382.
 DR InterPro: IPR001362; Glyco_hydro_32; 1.
 DR Pfam: PF00251; Glyco_hydro_32; 1.
 KW Glycoprotein; Glycosidase; Hydrolase.
 FT NON-TER 169
 FT NON-TER 169
 SQ SEQUENCE 169 AA; 19268 MW; B74A1632CD3AFC0 CRC64;

Query Match 2.0%; Score 8; DB 10; Length 169;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 66 LPTGNPVI 73
 DB 52 LPTGNPVI 59
 RESULT 12
 091596 PRELIMINARY; PRT; 210 AA.
 ID 091596
 AC 091596;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE HIGH MOBILITY GROUP PROTEIN-1.
 GN HMG-1.
 OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 CC Xenopodinae; Xenopus.
 OX NCB1_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96174815; PubMed=859938;
 RA Knightingale K., Dimitrov S., Reeves R., Wolffe A.P.;
 RT "Evidence for a shared structural role for HMG1 and linker histones B4
 and H1 in organizing chromatin.";
 RL EMBO J. 15:548-561(1996).
 DR EMBL: U21933; AAC59859.1; -;
 DR HSP: P07155; IAA.
 DR InterPro: IPR000135; Highmobility_12.

DR InterPro: IPR000910; HMG_12_box.
 DR Pfam: PF00505; HMG_box; 2.
 DR PRINTS: PR00886; HIGHMOBILITY12.
 DR SMART: SM00398; HMG; 2.
 DR PROSITE: PS00353; HMG1_2; 1.
 KW Chromosomal protein; DNA-binding; Nuclear protein.
 SQ SEQUENCE 210 AA; 24442 MW; 192CD46D594FF447 CRC64;

Query Match 2.0%; Score 8; DB 13; Length 210;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 286 SDRPKIK 293
 DB 106 SDRPKIK 113

RESULT 13
 ID 09CKU9 PRELIMINARY; PRT; 227 AA.
 AC 09CKU9;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE GPM.
 GN GPM OR PM1506.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_Taxid=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006187; AK03590.1;
 DR InterPro: IPR01345; PG_mutase.
 DR Pfam: PR00300; PGAM; 1.
 KW Complete Proteome.
 SQ SEQUENCE 227 AA; 25954 MW; B0928CC77B5C6A3 CRC64;

Query Match 2.0%; Score 8; DB 2; Length 227;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 194 GRRVLYTA 201
 DB 173 GRRVLYTA 180

RESULT 14
 ID P70441 PRELIMINARY; PRT; 355 AA.
 AC P70441;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE PROTEIN CO-FACTOR.
 GN SLC9A3R1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57 BLACK/6; TISSUE=KIDNEY;
 RA Wehman E.J., Stebbins D.A., Shenolikar S.;
 RT Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U74079; AAB17569.1; -.

DR HSP; Q12923; 3PDZ.
 DR MGD; MG1:1349482; SLC9a3r1.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 SQ SEQUENCE 355 AA; 38600 MW; 331F6BEE31DA0A11 CRC64;

Query Match 2.0%; Score 8; DB 11; Length 355;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 152 ERLKLG 159
 DB 97 ERLKLG 104

RESULT 15
 ID 09X855 PRELIMINARY; PRT; 396 AA.
 AC 09X855;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE PUTATIVE ASPARTATE AMINOTRANSFERASE.
 GN SCH10.36.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_Taxid=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb streptomycetes coelicolor A3(2) chromosome.";
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrett B.G., Rajandream M.A.;
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach W., Kleser H.M., Denapalpe D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL049754; CAB42045.1;
 DR HSSP: Q56232; IBKG.
 DR InterPro: IPR001176; ACC-synthase.
 DR InterPro: IPR001511; Aminotran_1.
 DR Pfam: PF00155; aminotran_1; 1.
 DR PRINTS: PR00753; ACCSYNTHASE.
 KW Transferase; Aminotransferase.
 SQ SEQUENCE 396 AA; 42785 MW; 727DBF14C347EBDE CRC64;

Query Match 2.0%; Score 8; DB 2; Length 396;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 197 VLVYTAGT 204
 DB 99 VLVYTAGT 106

Search completed: January 31, 2002, 13:37:33

C.Keywords: hydrolase

Query Match

Best Local Similarity 14.1%; Score 22; DB 2; Length 154;
Pred. No. 3-4e-15;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GSFAMVDPGMDGNLTMLTYNAS 110

DB 89 GSFAMVDPGMDGNLTMLTYNAS 110

RESULT 3

DB4965

dihydroilipomide S-succinyltransferase (EC 2.3.1.61) [imported] - Buchnera sp. (strain A

C.Species: Buchnera sp.

C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C.Accession: D84965

R.Shigenobu, S.; Matanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A.Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A

A.Reference number: A84930; MUID:20445173

A.Accession: D84965

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-420 <STO>

A.Cross-references: GB:AP000398; GSPDB:GN0144

A.Experimental source: strain APS

C.Genetics:

A.Gene: SUCB: BM303

C.Superfamily: dihydroilipomide acetyltransferase; lipoyl/biotin-binding homology

C.Keywords: acyltransferase; coenzyme A

Query Match

Best Local Similarity 5.1%; Score 8; DB 2; Length 420;
Pred. No. 3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 EKEGKVTI 52

DB 62 EKEGKVTI 69

RESULT 4

DB4965

flagellar protein flis - Bacillus subtilis

C.Species: Bacillus subtilis

C.Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C.Accession: I40398; E69625

R.Chen, L.; Helmann, J.D.

J. Bacteriol. 176, 3093-3101, 1994

A.Title: The Bacillus subtilis sigma D-dependent operon encoding the flagellar proteins

A.Reference number: I40396; MUID:94252974

A.Accession: I40398

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-133 <RES>

A.Cross-references: EMBL:Z31376; NID:9499379; PIDN:CA83249.1; PID:9499382

R.Kumet, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Berter

C. Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Cho

A. Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabbre, C.; Ferrati, E.

Nature 390, 249-256, 1997

A.Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.

Koetter, P.; Koningsstein, G.; Krog, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A.Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanton

A.Authors: Scheich, S.; Schroeter, R.; Scoffone, F.; Sekiyuchi, J.; Sekowska, A.; Seron

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yaguchi, K.; Yata, K.; Yoshida, K

A.Authors: Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Danchin, A.

Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A.Reference number: A69580; MUID:98044033

A.Accession: E69625

A.Status: nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-133 <KUN>

A.Cross-references: GB:299122; GB:AL009126; NID:92636029; PIDN:CA815550.1; PID:926360

A.Experimental source: strain 168

C.Genetics:

A.Gene: flis

C.Superfamily: flagellar protein flis

Query Match

Best Local Similarity 4.5%; Score 7; DB 1; Length 133;
Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LTLMLYN 108

DB 23 LTLMLYN 29

RESULT 5

DB4101

flagellar protein flis [imported] - Bacillus halodurans (strain C-125).

C.Species: Bacillus halodurans

C.Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

C.Accession: E84101

R.Takami, H.; Nakasone, K.; Takaki, Y.; Mieno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A.Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A.Reference number: A83650; MUID:20263314

A.Accession: E84101

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-133 <STO>

A.Cross-references: GB:AP001519; GB:BA000004; NID:910176109; PIDN:BA807332.1; GSPDB:G

A.Experimental source: strain C-125

C.Genetics:

A.Gene: flis

C.Superfamily: flagellar protein flis

Query Match

Best Local Similarity 4.5%; Score 7; DB 2; Length 133;
Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LTLMLYN 108

DB 23 LTLMLYN 29

RESULT 6

DB4101

all protein precursor - Yersinia enterocolitica

C.Species: Yersinia enterocolitica

C.Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 02-Sep-2000

C.Accession: A35123

R.Miller, V.L.; Biltska, J.B.; Falkow, S.

J. Bacteriol. 172, 1062-1069, 1990

A.Title: Nucleotide sequence of the Yersinia enterocolitica AII gene and characteriza

A.Reference number: A35123; MUID:90130261

A.Accession: A35123

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-178 <MTI>

A.Cross-references: GB:M29945; NID:9155437; PIDN:AAA86694.1; PID:9155438

C.Superfamily: phage lambda membrane protein Iom

Query Match

Best Local Similarity 4.5%; Score 7; DB 2; Length 178;
Pred. No. 16;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 GVIGSFA 92
|||||||
DB 65 GVIGSFA 71

RESULT 7

S74033
alkyl hydroperoxide reductase homolog - Sulfolobus solfataricus

A:Alternate names: protein c0215
C:Species: Sulfolobus solfataricus
C>Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000
C:Accession: S74033
R:Seisen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.
Mol. Microbiol. 22, 175-191, 1996
A:Title: Organizational characteristics and information content of an archaeal genome: I
A:Reference number: S73076; MUID:97055432
A:Accession: S74033
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-215 <SEN>
A:Cross-references: EMBL:Y08256; NID:g1707679; PIDN:CAA69447.1; PID:g1707727
A:Experimental source: strain P2
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996
C:Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 prot

Query Match 4.5%; Score 7; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 RIKLPDD 71
|||||||
DB 24 RIKLPDD 30

RESULT 8

C64491
phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase - Methanococcus

C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: C64491
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: C64491
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-237 <80U>
A:Cross-references: GB:U67594; GB:L77117; NID:g1592160; PIDN:AAB99553.1; PID:g1592163; T
C:Genetics:
A:Map position: FOR150884-1509597
C:Superfamily: N-(5'-phospho-D-ribosylformimino)-5-amino-1-(5''-phosphoribosyl)-4-imidaz

Query Match 4.5%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 KKKKVI 52
|||||||
DB 131 KKKKVI 137

RESULT 9

A69000
conserved hypothetical protein MTH1 - Methanobacterium thermoautotrophicum (strain Delta
C:Species: Methanobacterium thermoautotrophicum
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A69000

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Mierzdowski, J.; Gibson, R.; Jivani, K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: A69000
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <MTH>
A:Cross-references: GB:A6000795; GB:A6000666; NID:g2621036; PIDN:AAB84510.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1
C:Superfamily: conserved hypothetical protein MTH1

Query Match 4.5%; Score 7; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 KELLRP 16
|||||||
DB 149 KELLRP 155

RESULT 10

A64149
hypothetical protein HI0345 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C:Accession: A64149
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
Kerlavage, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: A64149
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-279 <TIGR>
A:Cross-references: GB:U32719; GB:I42023; NID:g1573310; PIDN:AAC22006.1; PID:g1573315
A>Note: Best homolog was a hypothetical protein from Escherichia coli
C:Superfamily: conserved hypothetical protein HI0345; ferredoxin 214Fe-4S) homology
F:65-130/Domain: ferredoxin 214Fe-4S) homology <FER6>

Query Match 4.5%; Score 7; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 SLAREGV 87
|||||||
DB 48 SLAREGV 54

RESULT 11

A70435
cysteine synthase (EC 4.2.99.8) cysM [similarity] - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Jun-2000
C:Accession: A70435
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70435
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-327 <AOF>

01-NOV-1999 (TREMblrel. 12, last annotation update)
 DE K08C9.6 PROTEIN.
 GN Caenorhabditis elegans.
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NCBI_TaxID=6239;
 OX NCBI_TaxID=6239;
 RN SEQUENCE FROM N.A.
 RP Lennard N.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lighting J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.,
 RA "2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RT Nature 368:32-38(1994).
 RL EMBL; 281567; CAB04589.1;
 DR SEQUENCE 127 AA; 14638 MW; B169A1E75006C2CE CRC64;
 SQ

Query Match 2.0%; Score 8; DB 5; Length 127;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 248 FRIRIKL 255
 Db 85 FRIRIKL 92

RESULT 10
 O9RIAL PRELIMINARY; PRT; 142 AA.
 AC 09RIAL;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE SODIUM-HYDROGEN EXCHANGER REGULATORY FACTOR (FRAGMENT).
 GN SLC9A3R1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SVJ.
 RA Weinman E.J., Steplock D.A., Zhang X., Akhter S., Shenolikar S.;
 RT "Molecular cloning of the cDNA and promoter sequences for the mouse
 sodium-hydrogen exchanger regulatory factor."
 RL Biochim Biophys. Acta 0:0-0(1999).
 DR EMBL; AF154812; MAD49224.1;
 DR HSSP; Q12923; 3PDZ.
 DR MGD; MGI:1349482; SLC9A3R1.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 FT NON_TER 142 142
 SQ SEQUENCE 142 AA; 15376 MW; A8994D6A865B283A CRC64;
 Match 2.0%; Score 8; DB 11; Length 142;
 Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 152 ERLKGLV 159
 Db 97 ERLKGLV 104

RESULT 11
 O9ZSW7 PRELIMINARY; PRT; 169 AA.
 AC 09ZSW7;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE BETA-FRUCTOFURANOSIDASE (EC 3.2.1.26) (FRAGMENT).
 OS Hamamelis virginiana.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 OC Saxifragales; Hamamelidaceae; Hamamelis.
 NCBI_TaxID=4397;
 OX NCBI_TaxID=4397;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LEAF;
 RA "Isolation of Invertase gene sequences from witch hazel (Hamamelis
 virginiana)."
 RT Thesis (1998), Pennsylvania State University.
 RL -1- CATALYTIC ACTIVITY: HYDROLISTS OF TERMINAL NON-REDUCING BETA-D-
 CC FRUCTOFURANOSIDE RESIDUES IN BETA-D-FRUCTOFURANOSIDES.
 CC -1- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
 DR EMBL; AF091547; AAC99430.1;
 DR Mendel; 36382; Hamv1:1002:36382.
 DR InterPro; IPR001362; Glyco_hydro_32.
 DR Pfam; PF00251; Glyco_hydro_32; 1.
 DR Glycoprotein; Glycosidase; Hydrolase.
 FT NON_TER 1 1
 SQ SEQUENCE 169 AA; 19268 MW; B74A1632C0D3AFC0 CRC64;

Query Match 2.0%; Score 8; DB 10; Length 169;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 66 LPTGNPVI 73
 Db 52 LPTGNPVI 59
 RESULT 12
 O91596 PRELIMINARY; PRT; 210 AA.
 AC 091596;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE HIGH MOBILITY GROUP PROTEIN-1.
 GN HMG-1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96174815; PubMed=8599938;
 RA Knightingale K., Dimitrov S., Reeves R., Wolfe A.P.;
 RT "Evidence for a shared structural role for HMG1 and linker histones B4
 and H1 in organizing chromatin."
 RL EMBO J. 15:548-561(1996).
 DR EMBL; U21933; AAC59859.1;
 DR HSSP; P07135; 1AAB.
 DR InterPro; IPR000135; Highmobility_12.

InterPro: IPR000910; HMG_12_box.
 DR Pfam: PF00505; HMG_box; 2.
 DR PRINTS: PR00886; HIGHMOBLTY12.
 DR SMART: SM00398; HMG; 2.
 DR PROSITE: PS00353; HMG1_2; 1.
 DR Chromosomal protein; DNA-binding; Nuclear protein.
 SC SEQUENCE 210 AA; 24442 MW; 192CD46D694FF47 CRC64;

Query Match 2.0%; Score 8; DB 13; Length 210;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 286 SDFRPXIK 293
 DB 106 SDFRPXIK 113

RESULT 13
 ID Q9CKU9 PRELIMINARY; PRT; 227 AA.
 AC Q9CKU9;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE GPM. OR PM1506.
 GN Pasteurella multocida.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RA MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70."
 RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006187; AAK03590.1;
 DR InterPro: IPR001345; PG_mutase.
 DR Pfam: PF00300; PGAM; 1.
 KW Complete proteome.
 SC SEQUENCE 227 AA; 25954 MW; B0928CC77B5C6A3 CRC64;

Query Match 2.0%; Score 8; DB 2; Length 227;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 194 GRRVLVTA 201
 DB 173 GRRVLVTA 180

RESULT 14
 ID P70441 PRELIMINARY; PRT; 355 AA.
 AC P70441;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PROTEIN CO-FACTOR.
 GN SLC9A3R1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57 BLACK/6; TISSUE=KIDNEY;
 RA Weinman E.J., Steplock D.A., Shenolikar S.;
 RT Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U74079; AAB17569.1;

HSP; Q12923; 3PDZ.
 DR GSD; MG1:1349482; SLC9A3r1.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 SC SEQUENCE 355 AA; 38600 MW; 331F6BEE31DA0A11 CRC64;

Query Match 2.0%; Score 8; DB 11; Length 355;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 152 ERLKRLGV 159
 DB 97 ERLKRLGV 104

RESULT 15
 ID Q9X855 PRELIMINARY; PRT; 396 AA.
 AC Q9X855;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE ASPARTATE AMINOTRANSFERASE.
 GN SCH10_36.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomyces; Streptomyces; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RT Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.D., Parkhill J., Barrell B.G., Rastam M.A.;
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RT MOL. Microbiol. 21:77-96(1996).
 DR EMBL: AL049754; CAB42045.1;
 DR HSP; O56232; 1BKG.
 DR InterPro: IPR001176; ACC_synthase.
 DR InterPro: IPR001511; Aminotran_1.
 DR Pfam: PF00185; aminotran_1; 1.
 DR PRINTS: PR00753; ACCSYNTHASE.
 KW transferase; Aminotransferase.
 SC SEQUENCE 396 AA; 42785 MW; 727DBF14C347EBDE CRC64;

Query Match 2.0%; Score 8; DB 2; Length 396;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 197 VLVYAGAT 204
 DB 99 VLVYAGAT 106

Search completed: January 31, 2002, 13:37:33

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:35 ; Search time 130.99 seconds
(without alignments)
174.200 Million cell updates/sec

Title: US-08-957-709-71

Sequence: 1 MLPPMKIRKELIEPFSEE.....PYRGNVQSTRLAFSKRKKL 156

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size: 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

SPTREMBL_17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	5.1	459	5	Q26001	Q26001 plasmodium
2	7	4.5	34	4	Q9U0N9	Q9U0N9 homo sapien
3	7	4.5	37	8	Q9X005	Q9X005 toxoplasma
4	7	4.5	84	5	Q9N0C8	Q9N0C8 mesobuthus
5	7	4.5	84	5	Q9N682	Q9N682 mesobuthus
6	7	4.5	133	2	Q9K6M1	Q9K6M1 bacillus ha
7	7	4.5	141	2	Q9FCY3	Q9FCY3 yersinia en
8	7	4.5	175	4	Q9UHT7	Q9UHT7 homo sapien
9	7	4.5	180	5	Q9VIT0	Q9VIT0 drosophila
10	7	4.5	198	10	Q9FMX7	Q9FMX7 arabidopsis
11	7	4.5	250	2	Q9CKU7	Q9CKU7 pasteurilla
12	7	4.5	268	1	Q26109	Q26109 methanobact
13	7	4.5	305	11	Q9CVR9	Q9CVR9 mus musculu
14	7	4.5	362	2	Q9PIM7	Q9PIM7 campylobact
15	7	4.5	376	3	Q94671	Q94671 schizosach
16	7	4.5	387	2	Q9KE07	Q9KE07 bacillus ha
17	7	4.5	590	10	Q9FLH6	Q9FLH6 arabidopsis
18	7	4.5	616	12	P87708	P87708 fowlpox vir
19	7	4.5	661	1	Q30218	Q30218 archaeoglob

20	7	4.5	666	2	Q9KKR7	Q9KKR7 vibrio chol
21	7	4.5	671	3	Q9HEC3	Q9HEC3 neurospora
22	7	4.5	782	2	Q9PMT3	Q9PMT3 campylobact
23	7	4.5	874	10	Q92W97	Q92W97 arabidopsis
24	7	4.5	1084	5	Q9VAM3	Q9VAM3 drosophila
25	7	4.5	1151	4	Q9HCH6	Q9HCH6 homo sapien
26	7	4.5	1514	4	Q9HOB0	Q9HOB0 homo sapien
27	7	4.5	1766	12	Q9J599	Q9J599 fowlpox vir
28	7	4.5	1921	5	Q01349	Q01349 drosophila
29	7	4.5	1921	5	Q9V7C9	Q9V7C9 drosophila
30	7	4.5	2005	5	Q9VFF5	Q9VFF5 drosophila
31	6	3.8	21	2	Q57485	Q57485 enterococci
32	6	3.8	35	2	Q50845	Q50845 borrelia bu
33	6	3.8	37	5	Q9BM33	Q9BM33 theistate al
34	6	3.8	37	5	Q9BM32	Q9BM32 theistate al
35	6	3.8	55	2	Q51017	Q51017 borrelia bu
36	6	3.8	62	2	Q9K797	Q9K797 bacillus ha
37	6	3.8	64	5	P91000	P91000 caenorhabdi
38	6	3.8	68	2	Q9PBP2	Q9PBP2 xylella fas
39	6	3.8	72	12	Q997E5	Q997E5 human immun
40	6	3.8	78	4	Q14661	Q14661 homo sapien
41	6	3.8	83	2	Q86300	Q86300 microcystis
42	6	3.8	83	2	Q86306	Q86306 microcystis
43	6	3.8	83	2	Q86307	Q86307 microcystis
44	6	3.8	83	2	Q86310	Q86310 microcystis
45	6	3.8	83	2	Q86375	Q86375 microcystis

ALIGNMENTS

RESULT 1
ID Q26001 PRELIMINARY; PRT; 459 AA.
AC Q26001;
DT 01-NOV-1996 (TREMUREL. 01, Created)
DT 01-NOV-1996 (TREMUREL. 01, Last sequence update)
DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
DE RABGDI PROTEIN.
GN RABGDI.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=97001683; PubMed=844675;
RA Attal G., Langsley G.;
RT "A Plasmodium falciparum homologue of a rab specific GDP dissociation inhibitor".
RL Mol. Biochem. Parasitol. 79:91-96(1996).
DR EMBL; X93166; CAA63653.1; .
DR HSSP; P21856; IGND.
DR InterPro; IPR002005; Rab_GDI_REP.
DR Pfam; PF00996; GDI_1.
DR PRINTS; PR00891; RABGDIREP.
SQ SEQUENCE 459 AA; 52311 MW; DEICCA47B124DBEA CRC64;

Query Match 5.1%; Score 8; DB 5; Length 459;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LTERIKL 68
Db 212 LTERIKL 219

RESULT 2
ID Q9U0N9 PRELIMINARY; PRT; 34 AA.
AC Q9U0N9;
DT 01-MAY-2000 (TREMUREL. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE LSTR3A PROTEIN (FRAGMENT).
GN LSTR3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:99299247; PubMed:10369878;
RA Gilley J., Fried M.;
RT "Extensive gene order differences within regions of conserved synteny
RT between the fugu and human genomes: implications for chromosomal
RT evolution and the cloning of disease genes.";
RL Hum. Mol. Genet. 8:1313-1320(1999).
DR EMBL: Y17458; CAB44351.1; -;
DR NON_TER 1 1
FT NON_TER 34 34
SQ SEQUENCE 34 AA: 4023 MW: 1F90377C28FA152D CRC64;

Query Match 4.5%; Score 7; DB 4; Length 34;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 EAFVKGK 40
Db 26 EAFVKGK 32

RESULT 3
O9X005 PRELIMINARY; PRT; 37 AA.
AC O9X005;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE RIBOSOMAL PROTEIN L36.
OS Toxoplasma gondii.
OC Chloroplast.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae;
OC Toxoplasma.
OX NCBI_TaxID=5811;
RN [1]
RP SEQUENCE FROM N.A.
RA Kissinger J.C., Donald R.G., Moulton A.L., Gutell R., Aiello D.P.,
RA Lang-Unnasch N., Roos D.S.;
RT "Mapping, cloning, and complete sequence annotation of the 35-kb
RT plasmid genome of Toxoplasma gondii.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: U87145; AAD41142.1; -;
DR InterPro: IPR000473; Ribosomal_L36.
DR Pfam: PF00444; Ribosomal_L36; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 37 AA: 4546 MW: EDD8C59BEEB5E6AD CRC64;

Query Match 4.5%; Score 7; DB 8; Length 37;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 76 MKIRSSL 82
Db 1 MKIRSSL 7

RESULT 4
O9NJC8 PRELIMINARY; PRT; 84 AA.
AC O9NJC8;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE PUTATIVE NEUROTOXIN TX13.
OS Mesobuthus martensii (Menchurian scorpion) (Butus martensii).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC Butioidae; Butidae; Butus.
OX NCBI_TaxID=34649;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-VENOM GLAND;
RA Zhu S., Li W.;
RT "The full-length cDNA encoding a putative mammalian neurotoxin BmTx13
RT from Bm.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF15363; AAF31477.1; -;
DR InterPro: IPR003614; Kntcl.
DR InterPro: IPR002061; Scorpion_toxin.
DR Pfam: PF00537; toxin_3; 1.
DR PRINTS: PR00284; TOXIN.
DR Prodom: PD000908; Scorpion_toxin; 1.
DR SMART: SM00505; Kntcl; 1.
KW Neurotoxin.
SQ SEQUENCE 84 AA: 9319 MW: 467FFD159A223530 CRC64;

Query Match 4.5%; Score 7; DB 5; Length 84;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 66 IKLPDDV 72
Db 68 IKLPDDV 74

RESULT 5
ID O9N682 PRELIMINARY; PRT; 84 AA.
AC O9N682;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE NEUROTOXIN BMK11 PRECURSOR (TOXIN ALPHA-TX14).
OS Mesobuthus martensii (Menchurian scorpion) (Butus martensii).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC Butioidae; Butidae; Butus.
OX NCBI_TaxID=34649;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-VENOM GLAND CYTOPLASM;
RA Zeng X.-C., Zhu S.-Y., Li W.-X., Jiang D.-H.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE-VENOM GLAND;
RA Li W., Zhu S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA TISSUE-VENOM GLAND;
RA MEDLINE:20317244; PubMed:10858508;
RA Zhu S.-Y., Li W.-X., Zeng X.-C., Liu H., Jiang D.-H., Mao X.;
RT "Nine novel precursors of Butus martensii scorpion alpha-toxin
RT homologues.";
RL Toxicon 38:1653-1661(2000).
DR EMBL: AF114025; AAF34872.1; -;
DR EMBL: AF156169; AAF29462.1; -;
DR EMBL: AF156595; AAK06987.1; -;
DR InterPro: IPR003614; Kntcl.
DR InterPro: IPR002061; Scorpion_toxin.
DR InterPro: IPR001219; toxin.
DR Pfam: PF00537; toxin_3; 1.
DR PRINTS: PR00284; TOXIN.
DR Prodom: PD000908; Scorpion_toxin; 1.

DR SMART; SM00505; Knot1. 1.
 KW Signal; Neurotoxin. 19 POTENTIAL.
 FT CHAIN 20 84 NEUROTOXIN BMK11.
 FT SEQUENCE 84 AA; 9308 MW; C08EA4859A30B7B6 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 5; Length 84;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 66 IKLPDDV 72
 |||||
 Db 68 IKLPDDV 74

RESULT 6
 ID 09K6W1 PRELIMINARY; PRT; 133 AA.
 AC 09K6W1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE FLAGELLAR PROTEIN.
 GN FLIS OR BH3613.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86655;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AF001519; BAB07332.1; -;
 DR Interpro; IPR003713; FLIS.
 DR Pfam; PF02561; FLIS; 1.
 KW Complete proteome.
 SQ SEQUENCE 133 AA; 15241 MW; F6BB8C320D83F8A CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 2; Length 133;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 LTIMLYN 108
 |||||
 Db 23 LTIMLYN 29

RESULT 7
 ID 09FCY3 PRELIMINARY; PRT; 141 AA.
 AC 09FCY3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE ATTACHMENT INVASION LOCUS PROTEIN (FRAGMENT).
 GN AIL.
 OS Yersinia enterocolitica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gierczynski R., Jagielski M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF004311; AAF97354.1; -;

DR InterPro; IPR000758; Enterovir_OMP.
 DR PRINTS; PR00316; ENTEROVIR_OMP.
 DR PROSITE; PS00694; ENT_VIR_OMP_1; 1.
 FT NON_TER 1 1
 FT NON_TER 141 141
 SQ SEQUENCE 141 AA; 15593 MW; 4B689FPC3B13D8F8 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 2; Length 141;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 GVIGSFA 92
 |||||
 Db 46 GVIGSFA 52

RESULT 8
 ID 09UHT7 PRELIMINARY; PRT; 175 AA.
 AC 09UHT7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE PRO1779.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Bi J., Zhang Y., Liu M.,
 RA He F.;
 RT "Functional prediction of the coding sequences of 33 new genes deduced
 RT by analysis of cDNA clones from human fetal liver.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF118076; AAF22020.1; -;
 SQ SEQUENCE 175 AA; 20393 MW; 7F1365992AD04B01 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 4; Length 175;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 44 VEKEGV 50
 |||||
 Db 85 VEKEGV 91

RESULT 9
 ID 09VIT0 PRELIMINARY; PRT; 180 AA.
 AC 09VIT0;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE CG9337. PROTEIN.
 GN CG9337.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang O., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abdl J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 Borikova D., Botchan M.A., Bouck J., Brokstein P., Brothier P.,
 Butulis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Mishina N.V., Mobarry B., Murphy L., Muzny D.M., Nelson D.L.,
 Mount S.M., Moy M., Murphy B., Murray C., Morris J., Moshrefi A.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 "The genome sequence of *Drosophila melanogaster*."
 Science 287:2185-2195(2000).
 EMBL: AE003668. AAF53639.1.
 DR FlyBase: FB00032898; CG9337.
 DR InterPro: IPR000958; K4
 SO SEQUENCE 180 AA; 20170 MW; 8DFC247F100E92A4 CRC64;

Query Match 4.5%; Score 7; DB 5; Length 180;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 16 PFSEBSL 22
 Db 72 PFSEBSL 78

RESULT 10
 O9FMX7 PRELIMINARY; PRT; 198 AA.
 AC O9FMX7;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE GENOMIC DNA, CHROMOSOME 5, P1 CLONE:MKD15.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RX MEDLINE=98162728; PubMed=9501997;
 RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
 Takata S.;
 "Structural analysis of Arabidopsis thaliana chromosome 5. III.
 RT Sequence features of the regions of 1,191,918 bp covered by seventeen
 RT physically assigned pl clones."
 DL DNA Res. 4:401-414(1997).
 DR EMBL: AB007648; BAB1178.1; -
 DR InterPro: IPR000868; Isochorismatase.

DR Pfam: PF00857; Isochorismatase; 1.
 SO SEQUENCE 198 AA; 22318 MW; F0A53BC468F208A1 CRC64;
 Query Match 4.5%; Score 7; DB 10; Length 198;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 33 REAFVKG 39
 Db 147 REAFVKG 153

RESULT 11
 O9CKL7 PRELIMINARY; PRT; 250 AA.
 AC O9CKL7;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NABG.
 GN NABG OR PM1595.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella
 NCBI_TaxID=747;
 RX MEDLINE=21142866; PubMed=11248100;
 RC STRAIN=PM70;
 RA May B.-J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
 "Complete genome sequence of *Pasteurella multocida* PM70."
 Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006187; AAK03679.1;
 DR InterPro: IPR001450; 4FE4S_Ferredoxin.
 DR Pfam: PF00037; ferf, 2.
 DR PROSITE: PS00196; 4FE4S_FERREDOXIN; 1.
 KW Complete proteome.
 SO SEQUENCE 250 AA; 27277 MW; ED69B44B7CCD8A4 CRC64;

Query Match 4.5%; Score 7; DB 2; Length 250;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 81 SLAREGV 87
 Db 38 SLAREGV 44

RESULT 12
 O26109 PRELIMINARY; PRT; 268 AA.
 AC O26109;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CONSERVED PROTEIN.
 GN MTH1.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanothermobacter.
 NCBI_TaxID=145262;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doncic-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
 Aldredge T., Baschirzadeh R., Blakely D., Cook R., Gilbert K.,
 Harrison D., Hoang L., Kaegi P., Lamm W., Pollier B., Qi D.,
 Spadafora R., Vance R., Wang Y., Wierzbowski J., Gibson R.,
 Jivani N., Caruso A., Bush D., Safer H., Patel J.D., Prabhakar S.,
 McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,

RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltapH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 DR EMBL: AE000795; AAB84510.1; -;
 DR InterPro: IPR003750; DUF171.
 DR Pfam: PF02598; DUF171; 1.
 KW Complete proteome.
 SQ SEQUENCE 268 AA; 30191 MW; 1B97DBICEE27BAED CRC64;

Query Match 4.5%; Score 7; DB 1; Length 268;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 KEILIEP 16
 DB 149 KEILIEP 155

RESULT 13
 ID 09CVR9 PRELIMINARY; PRT; 305 AA.
 AC 09CVR9;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 1700055M05BRIK PROTEIN (FRAGMENT).
 GN 1700055M05BRIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi S., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasikawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glisli C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK006800; BAB24746.1; -;
 DR MGD: MGI:1921520; 1700055M05BRIK.
 DR InterPro: IPR001394; UCH-2.
 DR Pfam: PF00443; UCH-2; 1.
 DR PROSITE: PS00973; UCH_2_2; 1.
 DR PROSITE: PS50235; UCH_2_3; 1.
 FT NON_TER 1
 SQ SEQUENCE 305 AA; 34632 MW; 343B53F1A9B5F482 CRC64;

Query Match 4.5%; Score 7; DB 11; Length 305;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 34 EAFVKGK 40
 11111111

DB 220 EAFVKGK 226

RESULT 14
 ID 09PIW7 PRELIMINARY; PRT; 362 AA.
 AC 09PIW7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE TRANSMEMBRANE PROTEIN.
 GN CJO268C.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter
 OX NCBI_TaxID=197;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S.,
 RA Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL: AL139074; CAB72736.1; -;
 DR InterPro: IPR001107; Band_7.
 DR InterPro: IPR000163; Prohibitin.
 DR Pfam: PF01145; Band 7; 1.
 DR PRIMS: PR00679; PROHIBITIN.
 DR SMART: SM00244; PHB; 1.
 KW Complete proteome.
 SQ SEQUENCE 362 AA; 40260 MW; 16F5908969F29238 CRC64;

Query Match 4.5%; Score 7; DB 2; Length 362;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 NEPELR 117
 DB 217 NEPELR 223

RESULT 15
 ID 094671 PRELIMINARY; PRT; 376 AA.
 AC 094671;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PROBABLE HOMOSERINE DEHYDROGENASE (EC 1.1.1.3) (HDH).
 GN SPC0776.03.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Lyne M., Rajandream M.A., Barrell B.G., Wedler H., Kutzner M.,
 RA Wambolt R.;
 RT Submitted (JAN-1999) to the EMBL/Genbank/DBD databases.
 CC -1- CATALYTIC ACTIVITY: L-HOMOSERINE + NAD(P)(+) = L-ASPARTATE BETA-
 CC SEMIALDEHYDE + NAD(P)H.
 CC -1- PATHWAY: THIRD STEP IN CONVERSION OF L-ASPARTATE TO HOMOSERINE.
 CC HOMOSERINE PARTICIPATES IN THE BIOSYNTHESIS OF THRONINE AND THEN
 CC ISOLEUCINE AND IN THE BIOSYNTHESIS OF METHIONINE.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

Thu Jan 31 13:32:31 2002

us-08-957-709-71.rspt

Page 6

CC -1- SIMILARITY: TO OTHER SPECIES HOMOSERINE DEHYDROGENASE.
DR EMBL; AL035263; CAA22876.1; -.
DR InterPro; IPR001342; Homoserine_dh.
DR Pfam; PF00742; Homoserine_dh; 1.
DR PROSITE; PS01042; HOMOSER_DHCENASE; 1.
KW Oxidoreductase; NADP; Threonine biosynthesis; Isoleucine biosynthesis;
KW Methionine biosynthesis.
FT NP_BIND 13 20 NADP (POTENTIAL).
SQ SEQUENCE 376 AA; 40037 MW; 3A2BC3A2CBA9263C CRC64;

Query Match 4.5%; Score 7; DB 3; Length 376;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 45 EKEGRV 51
|||||
Db 287 EKEGRV 293

Search completed: January 31, 2002, 13:37:37
Job time: 162 sec


```

Db      65 GKALKFGGNGVEYETIDIEHILL-----YNECCOLLTPATANTISINIGIADNIYN 120
QY      120 YVYTTAFPHIPIIMIAPIAMETMYRHRIVENTIERLK-KLGVETIGRIEGRKAVASIDE 178
Db      121 TLAIMEFGKKPFIVPAMENMEN--AIRRHIDKLEKKRIYIISKFEGRKAKVANIED 178
QY      179 IYRYRYKAL-HKKTLEGRVLYTAGTREVYIDPIRITVASSGKMGVALAEADPRG-AV 236
Db      179 VYKAYETKGNLKEGNNVLLNGSTVEIDKVRYSILSSGKMGVALAEAFCKEGFYV 238
QY      237 TLIRKGVAKFRIRIKIKLVETVEEMLS-AIENELSRKKYDVYVMAAASDFPRKAE 295
Db      239 EVITAMGLPPYYIK--NHRVLTAKEMLKAE--LAKDFIITISSAIDPTVE-SFE 292
QY      296 GKIKSGRSTIELVPXNPKIIDRIKEIQPNVFLVYGRAB--TSKEKLEEGKRQIERAKA 353
Db      293 GRLSSEELILKL-KNRPVLEELRIRYKDYIIGFKAEYNLDEKELIRAKERLKNYML 351
QY      354 DLVVGNTLEA--FGSEENOVVLIGRDFTK-ELPKM--KKRELAERIMDEIEK 400
Db      352 NMIIANDLSKHFGDDYIEVYII---TKYEVEKISGSKREISERIVERVK 399

```

RESULT 2

```

DEF_BORBU 2
ID DEF_BORBU STANDARD: PRT: 390 AA.
AC 051752;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG.
DFF OR B80812.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
[1]
RN RP
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Castens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lattigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D., Hanson M.,
RA Peterson J., Keliavage A.R., Quackenbush J., Salzberg S., Hannon M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Gaitanar S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi."
RT Nucleotide 390:580-586(1997).
RT BUREDOFIERI.
CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE METABOLISM (BY SIMILARITY).
CC CC
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC CC
CC CC EMBL: AE001179; AAC67145.1; -
CC CC TIGR: B80812; -
CC CC InterPro: IPR003382; Flavoprotein.
CC CC Pfam: PF02441; Complete proteome.
CC CC KX Flavoprotein. Complete proteome.
CC CC SEQUENCE 390 AA; 44144 MW; 66868FAB172DE300 CRC64;

```

Query Match 27.2%; Score 543.5; DB 1: Length 390;
 Best Local Similarity 36.3%; Pred. No. 1.3e-26;
 Matches 144; Conservative 81; Mismatches 135; Indels 37; Gaps 14;

```

QY      19 KTIYXXPGSIAL-DYKACEGLIRHGAEVHAVMSAARTIHPYAMNLPNGPVYTEIT 77
Db      5 KHLIGTCGGIASYSVSVYSSVLKGRVYKVMYMONAKFTPLPLETISKNKTIITNM 64
QY      78 GR----IEVEVAGEHENKADLLVCPATANTISKIACIDPTVYVYTTAFPHIPIMI 133
Db      65 DLIDNEVEHIKIA---KWAHLILVPAVYNTISKIAGIADALTLTIIAS--TAPYF 118
QY      134 APAMETMYRHRIVRENIERLKLGVETIGPRIEGS-----RAKVASIDEIYRVIK 185
Db      119 ATAMNNIMYSNPLKENIKLTKTYNKFTEP--DKGFLACSSNALGRKLNEDKRIKIILN 176
QY      186 KLHKKT-LEGRVLYTAGTREVYIDPIRITVASSGKMGVALAEADPRG-VTLI--R 240
Db      177 EFNOKDYLNKKKILITASTPEELIDPIRYFSMTSGCKMFCLOAEVYKLGAVTTITGPT 236
QY      241 TGSYKAFRIKRIKIKLVETVEEMLS-AIENELSRKKYDVYVMAAASDFPRKIKAGKIKS 300
Db      237 NENDEGVNIIKIKTAMEMYKEALKIY-----NKFETIIGAAVAADFPRKHIIFNSKIK 290
QY      301 GR-STITIELVPXNPKIIDRIKEIQ-PNVFLVGFKAETSKREKLEEGKRQIERAKADLV 357
Db      291 NKINRLYIKLV-KNPDIIOHGNKLNQIVIGFCANESK-NLIGKAKKELTKKNLDFII 348
QY      358 GNTLEAFGESENOVVLIGRDFTEKELPKMKRRELAERI 394
Db      349 ANELKYGSKLKNVYIINKOSIKELPEMEKSEVAKEI 385

```

RESULT 3

```

DEF_HAEIN 3
ID DEF_HAEIN STANDARD: PRT: 400 AA.
AC P44953;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG.
DFF OR H10953.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
OX NCBI_TaxId=727;
[1]
RN RP
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Keliavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Keliavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uteerback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brendon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Georghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
RT Science 269:496-512(1995).
RT RT
CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE METABOLISM (BY SIMILARITY).
CC CC
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC CC
CC CC EMBL: U03276; AAC22614.1; -
CC CC TIGR: H10953; -
CC CC InterPro: IPR003382; Flavoprotein.

```

DR Pfam: PF02441; Flavoprotein; 1.
 KM Flavoprotein; Complete proteome.
 SQ SEQUENCE 400 AA; 43401 MW; 1D6FEAE20C26B913 CRC64;

Query Match 23.0%; Score 459.5; DB 1; Length 400;

Best Local Similarity 33.7%; Pred. No. 1,9e-21;
 Matches 137; Conservative 84; Mismatches 148; Indels 37; Gaps 17;

15 KLVGKTKYXXPGSIALDVKACGLRHG-AEYHAYMSEATYIIHPYANMLPTGNDVI 73
 2 KLNKQHIYVGTGIAAYKTEILRLRKAEYAVVLTAPAAEFVPLTLQATISGNAVS 61
 74 TEIIG-----FIEVELAGEHENKADLLVCPATANTISKACGIDDPVTYVTTAFPH 128
 62 QSLDDPOAEIAMGHIELA---KWADAIITAPASADFAFLTITGMANDLSTICLAT--N 115
 129 IPIIAPAMHETMYRHPVRENIRLKLGEVETGPRIT-----EEGRAKVASIDEIYR 182
 116 APIPLAPAMNQMTHOSTIQONLTTLQTRGIELIGPNSGFACGDMGKGRMSEPEET-FT 174
 183 VIRKL--HKRTLEGKRVLTAGATREYIDPIRFTTNASSSGKMGVALAEADFRGA-VTLI 239
 175 ALSDFESQKODLGUNVITAGPTREAIIDPVYRISNHSKMGFAIAEAFKRGANVTLLI 234
 240 RTKGSYAFRIKRI-KLKVETVEEMLSA-IENELRSKKYDVVYMAAASDFPRKIKAEK 297
 235 --APVNLTPTRKNVNVINISAQEMQASLESAAVKNQ--FFICCAAAADRVETVEAOK 289
 298 I-KSGRSTIELVPPNPXPKIIDRIKEIOPN-VFLVGFKATSEKELIEGKQIERRAKADL 355
 290 IKKGGDEISIKLI-KNDIISDVGHILKTHRFYVGFAAET--QNVVDYAKKLEKKNLDM 346
 356 VVGVTL--EAFGSEENOVVILGRDFTKELPKMKKRELAERIMDEI 398
 347 ICANDVSGGVFNADENALQLEFRKNGHKILSKKVELADLVNEI 392

RESULT 4

DEF_SYNY3 STANDARD; PRT; 402 AA.

AC P73881;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG.
 GN DEF OR SL0250.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugijura M., Sasamoto S., Kimura T.,
 RA Hosonouchi T., Matsuno A., Muraki C., Wada T., Watanabe A.,
 RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC
 CC -I- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
 CC METABOLISM (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D90910; BAA17944.1; -
 DR InterPro: IPR003382; Flavoprotein.
 DR Pfam: PF02441; Flavoprotein; 1.
 KM Flavoprotein; Complete proteome.
 SQ SEQUENCE 402 AA; 43419 MW; 7E3F73DE185EC97C CRC64;

Query Match 22.0%; Score 439; DB 1; Length 402;

Best Local Similarity 32.0%; Pred. No. 3.4e-20;
 Matches 129; Conservative 78; Mismatches 160; Indels 36; Gaps 15;

18 GKIKVXXXXPGSIALDVKACB---GLIRHGAEVHAYMSEATYIIHPYANMLPTGNDVI 74
 4 CKRLILVGGGIAA-YKICEVVSQQLFQGGAEVAVIILAEKRYVTLPTLTTLARHAYG 61
 75 EITGF--IEHVELAGEHENKADLLVCPATANTISKACGI-DDTPYVTVTTAFPHPI 131
 62 DADWQPIHHRPLHIDGEMADIFLAPLTAHLAKIGHGADDLSTWVLASS---CPI 118
 132 MIAPAMHETMYRHPVRENIRLKLGEVETGPR-----IEGRAKVASIDEIYR 182
 119 LAPAMMTDMWEOBAVORNIQOL--LDORRYHLAPNGGLACDRGVGLAEPAQIMHR 176
 183 VIRKLH---KKTEGKRVLTAGATREYIDPIRFTTNASSSGKMGVALAEADFRGA-VTL 238
 177 LQALLFTGGQEDLGGKRLITLITAGSTQETLDVAVRFIGNPSTKMLALAQSGDGRVTL 236
 239 IRTGGSYAFRIKRIKLVETVEEMLSAIENELRSKKYDVVYMAAASDFPRKIKAEK 298
 237 IHGPIGLDPIPMGVISMAVNAEQALAQOQ--APLADWIVMAAADVAPQATWTGKI 294
 299 -KSGRSTIELVPPNPXPKIIDRIKEIOPNVFLVGFKATSEKELIEGKQIERRAKADLV 357
 295 AKQDLPDLAPVDDILAQVSKQKRPQQLVGFPAQAG--DIVTPAEKLDKRNLDVIV 352
 358 GNTLE---AFGSEENOVVILGRDFTK-ELPKMKKRELAERIM 395
 353 ANPIDQPSGFGTDSNQAVIIDRHGSQTALAPCSKILMAHRIM 395

RESULT 5

DEF_MYCTU STANDARD; PRT; 418 AA.

AC P71661;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG.
 GN DEF OR RV1391 OR MT1436 OR MTCY21B4.08.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=9825987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hohnsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Stalston J.E., Taylor K., Whitehead S., Barrett B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salberg S.L.,
RA DeJener A., Uteback T., Weidman J., Knout H., Gill J., Mikula A.,
RA Blisat W.,
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains".
CC Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -i- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z80108; CAB02174.1; -
DR EMBL: AE007015; AAK45701.1; -
DR TIGR: MT1436; -
DR TubercuList: RV1391; -
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
KW Flavoprotein; Complete proteome.
SQ SEQUENCE 418 AA; 43577 MW; 48BD95E536595506 CRC64;

Query Match 21.5%; Score 430.5; DB 1; Length 418;
Best Local Similarity 32.3%; Pred. No. 1,2e-19;
Matches 137; Conservative 78; Mismatches 160; Indels 49; Gaps 20;
OY 15 KLVGKRIYXXHSGSTAALDYKACEGLIRGAE---YNAVSEAKTKIHPYAMNLPNGN 70
DB 5 KIRPKQVIVGSGGIAA-YKACT-VKRLTASHRVRYPFESADRFVGAATFEALSGE 61
OY 71 PYITEI---TGTEIHEVLAGEHKNKADLLVCPATANTISKIACGIDTPTVTVTAAFP 127
DB 62 PCTGVFADVPVAPVHVG---GQADLVVAPATADLARAARADDLTATLLTA-- 115
OY 128 HIPIMAPMHEHYRHPYRENIERLKKLVEFTGPR---IEEGRKAVASIDEI-- 179
DB 116 RCPVLFAPAMHEMMLHAPYVDVNTLRRRGAVVLEPATGRLTGADSGGRIPAEAEITTT 175
OY 180 YVRVTKLHKR---TLEGRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA- 235
DB 176 LAQLLERDALPYDLAGSKLLVTAGTREPIDPVRFIGNRSSGQGVAVAVARQAD 235
OY 236 VTLI--RTGSGYKAFRIKIKLKVTEVEMLSAIENELRSKKYDVVIMAAVSDPRPKIK 293
DB 236 VLLIGHVAGLVDPAVEV--HVSAAOOLADAVSK--HAPPADVLVMAAVALADPRPAOV 291
OY 294 AEGKTKSG--RSITITELVPXNPKIIDRIKEIQ-----PNV-FLVGFKAET--SKKELIEE 343
DB 292 ATAATKKGVEGEPTEILL-RNDVDLAGVVARAHGOLPFRMAIVGFAAETGANGDVLFFH 350
OY 344 GKRQIERAKADLVGNLT---EAFGSENOVVLIGDFTKKE-LPMKKRRELAERIMDETE 399
DB 351 AAKILRRKGCDDLVAVNAVGEAFVSDNGLWLASDGTESALQSGSKTILMSRIVDAIV 410
OY 400 KKL 403
DB 411 TFLA 414

RESULT 6
DPP_ECOLI ID DPP_ECOLI STANDARD: PRT: 406 AA.
AC P24285; p76718;
DE 20-AUG-2001 (Rel. 21, Created)
DE 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN.
DE DPP OR B3639.

OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication".
RL Genomics 16:551-561(1993).
RN [2]
RP SEQUENCE OF 288-406 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=84057777; PubMed=6139280;
RA Lundberg L.G., Thoresson H.O., Karlstrom O.H., Nyman P.O.;
RT "Nucleotide sequence of the structural gene for dUTPase of
RT Escherichia coli K-12".
RL EMBL J. 2:967-971(1983).
CC -i- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L10328; AA61992.1; ALT_INIT.
DR EMBL: AE000441; AAC7663.1; ALT_INIT.
DR Ecogene: EG10004; dtp.
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
KW Flavoprotein; Complete proteome.
SQ SEQUENCE 406 AA; 43438 MW; CAD11B9347E8C6AB CRC64;

Query Match 21.2%; Score 424.5; DB 1; Length 406;
Best Local Similarity 33.4%; Pred. No. 2.7e-19;
Matches 138; Conservative 71; Mismatches 155; Indels 49; Gaps 18;
OY 16 LVGKRIYXXHSGSTAALDYKACEGLIR-HGAEVNNAVSEAKTKIHPYAMNLPNGNYIT 74
DB 3 LAGKRIVLGVSOGGIAAYKTPRELVRRLRDGADVRAVAMEAKAFETTPSLQAVSGYPSD 62
OY 75 EI-----TGTEIHEVLAGEHKNKADLLVCPATANTISKIACGIDTPTVTVTAAFPHI 129
DB 63 SLIDPAEAAMHIELG---KMDLVILAPATADLARAARADDLTATLLTA-- 115
OY 130 PIMIPAMHEHYRHPYRENIERLKKLVEFTGPR-----RIEGRKAVASIDEIIV 181
DB 117 PVAVLPAAMNOQVYRAAATOHNEVLASRGILLMCPDSGSAQCGIDGPR---MDPLTI 172
OY 182 RYIKKLH---KTEGKRVLTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VT 237
DB 173 VDMAVAHFSVNDLKHLMIMITAGTREPIDPVRFIGNSSGKMGFAIAAARAGAVT 232
OY 238 LIRFGSYKAFRIKIKLKVTEVEMLSAIENELRSKKYDVVIMAAVSDPRPKIKAGSK 297
DB 233 LVSGPVSLTPPEYK-RVDVMTALEMAAVNASVQOO--NIFIGAAVADRAAVALAEK 289
OY 298 IK-----SGRSTITELVPXNPKI---IDRIKETOPNVFLVGFKAETSKKLEEGKRIE- 349
DB 290 IKKATQGDDELTKIKV-KNPDIAGVAAALKDRP--YVGFRAETNN--VEETARQRI 343
OY 350 RAKADLVGNLT---EAFGSENOVVLIGDFTKELPKMKRRELAERIMDEI 398
DB 344 KKNLDLICANDVSQPTQGSNDNNALHLEWODGDKVLELKERELGOLLDEI 396


```

RESULT 7
DFF_STRMU STANDARD; PRT; 145 AA.
ID_DFF_STRMU
AC 054433;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DN/PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG (FRAGMENT).
GN DFF.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=NG8;
RX MEDLINE=96312359; PubMed=8763945;
RA Gutierrez J.A., Crowley P.J., Brown D.P., Hillman J.D.,
RA Youngman P., Bleiweis A.S.;
RA "Insertional mutagenesis and recovery of interrupted genes of
RT Streptococcus mutans by using transposon Tn917: preliminary
RT characterization of mutants displaying acid sensitivity and
RT nutritional requirements";
RL J. Bacteriol. 178:4166-4175(1996).
CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
CC METABOLISM (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U48885; AAC44502.1;
CC InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein; 1.
CC Flavoprotein.
CC FT NON_TER 1 145
CC SEQUENCE 145 AA; 15819 MW; 5920CDD57C8F2DAB CRC64;

```

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND INDUCTION.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20117504; PubMed=10652125;
RA Espinosa-Ruiz A., Bellis J.M., Serrano R., Cullanez-Macia F.A.;
RT "Arabidopsis thaliana AtHAL3: a flavoprotein related to salt and
RT osmotic tolerance and plant growth".
RL Plant J. 20:529-539(1999).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hutzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maltli R., Marzilli A.,
RA Maltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utechtack T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana".
RL Nature 408:816-820(2000).
CC -1- FUNCTION: INVOLVED IN PLANT GROWTH AND SALT AND OSMOTIC TOLERANCE.
CC MAY CATALYZE THE ALPHA,BETA-DEHYDROXYMETHYLATION OF A PEPTIDYL CYSTEINE
CC OF A CELLULAR PARTNER (BY SIMILARITY).
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, SHOOTS, LEAVES, FLOWERS,
CC DEVELOPING SILIQUES AND SEEDS.
CC -1- INDUCTION: BY SALT STRESS.
CC -1- SIMILARITY: SOME, TO YEAST AND C.TROPICALIS SIS2/HAL3 AND YEAST
CC YKL088W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U80192; AAB53106.1;
CC DR EMBL: AC020889; AAF79709.1; ALT. INT.
CC DR Mendel: 12633; Arabid:1833;12633.
CC InterPro: IPR003382; Flavoprotein.
CC Pfam: PF02441; Flavoprotein; 1.
CC Flavoprotein; FMN.
CC KW SEQUENCE 201 AA; 22415 MW; 7962655E408FA64C CRC64;

```

```

Query Match 10.08; Score 199; DB 1; Length 145;
Best Local Similarity 37.28; Pred. No. 5.1e-06;
Matches 55; Conservative 22; Mismatches 57; Indels 14; Gaps 4;

```

```

Query Match 8.18; Score 162; DB 1; Length 201;
Best Local Similarity 27.18; Pred. No. 0.0014;
Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;

```

OY 115 DDPVTVTTAFPHIPIMAPAMHETMRHPVIRENIEERLKKGVETGCP 164
 DB 111 DNLICITIRAMWDYKPLFVAPAMNTLMNNPTEERHLTLDLGLITLLP 160

RESULT 9
 HL3A_ARATH STANDARD: PRT: 209 AA.
 ID HL3A_ARATH
 AC Q9SWE5;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HALOTOLERANCE PROTEIN HAL3A.
 GN HAL3A OR AT3G18030 OR MBG14.2.
 OS Arabidopsis thaliana (mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CC NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND INDUCTION.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20117504; PubMed=10652125;
 RA Espinosa-Ruiz A., Belles J.M., Serrano R., Cullanez-Macia F.A.;
 RT "Arabidopsis thaliana ATHAL3: a flavoprotein related to salt and
 RT osmotic tolerance and plant growth."
 RL Plant J. 20:529-539(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20444410; PubMed=10986463;
 RA Albert A., Martinez-Ripoll M., Espinosa-Ruiz A., Yennush L.,
 RA Cullanez-Macia F.A., Serrano R.;
 RT "The X-ray structure of the FMN-binding protein ATHAL3 provides the
 RT structural basis for the activity of a regulatory subunit involved in
 RT signal transduction."
 RL Structure 8:961-969(2000).
 CC -1- FUNCTION: INVOLVED IN PLANT GROWTH AND SALT AND OSMOTIC TOLERANCE.
 CC MAY CATALYZE THE ALPHA,BETA-DEHYDROGENATION OF A PEPTIDYL CYSTEINE
 CC OF A CELLULAR PARTNER.
 CC -1- COFACTOR: FMN.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, SHOOTS, LEAVES, FLOWERS,
 CC DEVELOPING SILICLES AND SEEDS WITH HIGHEST EXPRESSION IN SEED
 CC EMBRYOS AND PHLOEM.
 CC -1- INDUCTION: BY SALT STRESS.
 CC -1- SIMILARITY: SOME, TO YEAST AND C.TROPICALIS SIS2/HAL3 AND YEAST
 CC YKL088W.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF166262; AAD51616.1; -;
 CC EMBL; AB026641; BAB01351.1; -;
 CC PDB; 1E20; 11-SEP-00.
 CC InterPro; IPR003382; Flavoprotein.
 CC Pfam; PF02441; Flavoprotein; 1.
 CC Flavoprotein; FNM; 3D-structure.
 CC KW SEQUENCE 209 AA; 23355 MW; 3AB1BB364F840DE CRC64;

Query Match 7.9%; Score 158; DB 1; Length 209;
 Best Local Similarity 25.0%; Pred. No. 0.0026;
 Matches 52; Conservative 44; Mismatches 68; Indels 44; Gaps 10;

OY 11 TKSRLVGKKIYXXPGSIALDVACGLIRHGAEVHVSSEA-----TKIHPEY-----62.
 DB 15 TTPRK---PRVLLASGVSAAIKFGNLCHCFTEMEVRAVYTKSLHFLDKSLPQEVTL 71
 OY 63 -----AMNLPGNPVYTEITGFIHEVELAGEHNNADLLVCPATNYSKAGCID 114
 DB 72 YTDEWSSWN-KIDPVL-----HIEL-----RRMADVLVIAPLSANTLKGKAGCIG 118
 OY 115 DDPVTVTTAFPHIPIMAPAMHETMRHPVIRENIEERLKKGVETGCP---RI---EE 168
 DB 119 DNLICITIRAMWDYKPLFVAPAMNTLMNNPTEERHLTLDLGLITLLPILKKRLACGDY 178
 OY 169 GRKAVASIDEIYRYIK-----KLHKT 191
 DB 179 GNGAMAE-PSLITSTVRLFWESQAHQOT 205

RESULT 10
 Y113_YEAST STANDARD: PRT: 365 AA.
 ID Y113_YEAST
 AC P40506;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HYPOHETICAL 41.9 KDA PROTEIN IN SDS-TSII INTERGENIC REGION.
 GN Y11083C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.
 CC NCBI_TaxID=4932;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Bartlett B.G., Badcock R., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentes S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 246728; CAAB6711.1; -;
 CC SGD; S0001345; Y11083C.
 CC KW Hypothetical protein.
 CC SEQUENCE 365 AA; 41893 MW; D67655A7A3DDDDAB CRC64;

Query Match 7.5%; Score 150.5; DB 1; Length 365;
 Best Local Similarity 23.2%; Pred. No. 0.015;
 Matches 77; Conservative 50; Mismatches 120; Indels 85; Gaps 14;

OY 127 PHIPIMAPAMHE--TMRHPVIRENIEERLKKGVETGCPRIEGRKAVASIDEIV--YR 182
 DB 2 PCLPVLNRPQIHTSVTEISHADRTIKELPVPVATTEEOQFKNPRAPIDELTKDAK 61
 OY 183 VIKKLHKKTEGKRVLYTAGATREYID--PIRFTNASSGKMGVALAEADFG-AVTLI 239
 DB 62 EFIDLYSLKRNKIVLITNSGTTVPLENNTVRFIDNFAGSRGASSABOFLANGSVITFL 121

DT 15-DEC-1998 (Rel. 37, last sequence update)
 DT 20-AUG-2001 (Rel. 40, last annotation update)
 DE CARBAMOYL-PHOSPHATE SYNTHASE [AMMONIA], MITOCHONDRIAL PRECURSOR
 DE (EC 6.3.4.16) (CARBAMOYL-PHOSPHATE SYNTHETASE I) (CPSASE I).
 GN CPSI.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=92084128; PubMed=1840546;
 RA Haraguchi Y., Uchino T., Takiguchi M., Endo F., Mori M.,
 RA Matsuda I.;
 RT "Cloning and sequence of a cDNA encoding human carbamyl phosphate
 RT synthetase I: molecular analysis of hyperammonemia.";
 RL Gene 107:335-340(1991).
 RN [2]
 RP SEQUENCE FROM N.A.; VARIANT CPSI DEF. MET-544, AND VARIANT ALA-344.
 RC TISSUE=Liver;
 RX MEDLINE=98375696; PubMed=9711878;
 RA Finch U., Kohlschütter A., Schaefer H., Sperhake K., Colombo J.-P.,
 RA Gal A.;
 RT "Prenatal diagnosis of carbamoyl phosphate synthetase I deficiency by
 RT identification of a missense mutation in CPSI.";
 RL Hum. Mutat. 12:206-211(1998).
 CC -1- FUNCTION: INVOLVED IN THE UREA CYCLE OF UREOTELIC ANIMALS WHERE
 CC THE ENZYME PLAYS AN IMPORTANT ROLE IN REMOVING EXCESS AMMONIA
 CC FROM THE CELL.
 CC -1- CATALYTIC ACTIVITY: 2 ATP + NH(3) + CO(2) + H(2)O = 2 ADP +
 CC ORTHOPHOSPHATE + CARBAMOYL PHOSPHATE.
 CC -1- ENZYME REGULATION: REQUIRES N-ACETYLGUTAMATE AS AN ALLOSTERIC
 CC ACTIVATOR.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
 CC -1- TISSUE SPECIFICITY: PRIMARY IN THE LIVER AND SMALL INTESTINE.
 CC -1- DISEASE: DEFECTS IN CPSI ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
 CC METABOLIC DISORDER THAT CAUSE A TYPE OF HYPERAMMONEMIA. CLINICAL
 CC SYMPTOMS ARE VOMITING IN INFANCY, PROTEIN INTOLERANCE,
 CC INTERMITTENT ATAXIA, SEIZURES, LETHARGY, AND MENTAL RETARDATION.
 CC -1- SIMILARITY: TO OTHER CARBAMOYL-PHOSPHATE SYNTHETASES. ALSO
 CC CONTAINS A GLUTAMINE AMIDOTRANSFERASE-LIKE DOMAIN BUT THE CYSTEINE
 CC RESIDUE WHICH IS ESSENTIAL FOR AMIDO TRANSFERASE ACTIVITY, IS
 CC MISSING IN THIS PROTEIN SO THIS DOMAIN IS NO LONGER ACTIVE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D90282; BAA14328.1; -;
 DR EMBL: Y15793; CAAT5785.1; -;
 DR PIR: J01348; J01348.
 DR HSSP: P00968; LUDB.
 DR MEROPS: M38.971; -;
 DR MIM: 237300; -;
 DR InterPro: IPR001317; CPS_GATase.
 DR InterPro: IPR000901; CPSase.
 DR InterPro: IPR002474; CPSase_sm_chain.
 DR InterPro: IPR000991; GATase_1.
 DR InterPro: IPR000384; MGS.
 DR Pfam: PF00289; CPSase_L_chain; 2.
 DR Pfam: PF00988; CPSase_sm_chain; 1.
 DR Pfam: PF02142; MGS; 1.
 DR PRINTS: PR00098; CPSASE.
 DR PRINTS: PR00099; CPSGATASE.
 DR PROSITE: PS00866; CPSASE_1; 2.
 DR PROSITE: PS00867; CPSASE_2; 2.
 KW ligase; Repeat; Transit peptide; Mitochondrion; ATP-binding;

KW	urea cycle; Polymorphism; Disease mutation.	TRANSIT	1	38	MITOCHONDRION (BY SIMILARITY).
FT	CHAIN	39	1500		CARBAMOYL-PHOSPHATE SYNTHASE [AMMONIA].
FT	NP_BIND	571	626		ATP (2).
FT	NP_BIND	718	768		ATP (1).
FT	NP_BIND	1113	1171		ATP (2).
FT	NP_BIND	1259	1302		ATP (1).
FT	DOMAIN	39	219		ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
FT	DOMAIN	220	410		HOMOLOG.
FT	DOMAIN	411	1500		GLUTAMINE AMIDOTRANSFERASE-LIKE.
FT	REPEAT	419	876		CARBAMOYL-PHOSPHATE SYNTHETASE LARGE
FT	REPEAT	970	1410		CHAIN.
FT	VARIANT	344	344		
FT	VARIANT	544	544		
FT	CONFLICT	111	111		T -> A.
FT	CONFLICT	279	279		/FTId=VAR_006834.
FT	CONFLICT	338	338		T -> M (IN CPSI DEFICIENCY).
FT	CONFLICT	718	722		/FTId=VAR_006835.
FT	CONFLICT	729	729		A -> S (IN REF. 1).
FT	CONFLICT	749	749		R -> Q (IN REF. 1).
FT	CONFLICT	1161	1162		G -> C (IN REF. 1).
FT	CONFLICT	1204	1205		RUSRS -> KMSPN (IN REF. 1).
FT	CONFLICT	1254	1254		A -> T (IN REF. 1).
FT	CONFLICT	1266	1266		E -> G (IN REF. 1).
FT	CONFLICT	1283	1283		EH -> AT (IN REF. 1).
FT	CONFLICT	1303	1303		GD -> EN (IN REF. 1).
FT	CONFLICT	1406	1406		I -> N (IN REF. 1).
FT	CONFLICT	1406	1406		F -> S (IN REF. 1).
FT	CONFLICT	1406	1406		M -> L (IN REF. 1).
FT	CONFLICT	1406	1406		A -> V (IN REF. 1).
FT	CONFLICT	1406	1406		T -> N (IN REF. 1).
SO	SEQUENCE	1500 AA;	164939 MW;	ES3A22D77563961D CRC64;	

Query Match	6.7%;	Score 134;	DB 1;	Length 1500;
Best Local Similarity	20.6%;	Pred. No. 0.83;		
Matches 93;	Conservative 63;	Mismatches 134;	Indels 162;	Gaps 22;
10 ATSRKLVGKKIYXXXPGSIALDV----	KACGLIRHGAEVAVNAVSEATKIIHYANN 65			
205 STADVAVYGR-----	GNPTKVAVDCGIRKNVIRLVKRGAEVHLV-----	PWN 248		
66 LP-----	TGNVITEITGFEIHEVLAEHKKADLIVCPATANTISKIAC 111			
249 HDPTKMEYDSIIILAGSPNALAE--	PLIQNVKIKIESDKKEPLFGI--STGLILIGLA 304			
112 GI-----	DDPTVTVT-----TAEPH-----	IPMIAP-----	A 136.	
305 GATYKSMANRGOQNPVINTKQAFITAONHGYALDNTLPAGMKRPLFVNVDQTNEGI 364				
137 MHTMTNRHPVRENIERLKKLVGEFICPRLEBRRAVASIDELVYVKKLKKTKTEGR 196				
365 MHES---KPF-----	FAVQF--HPEVTPGPIDTEYLDFESFLIK-----	GRA 403		
197 VLVT-----	GATREYIDPIRFITNASSGKMVALAEADRAVTLIRTSYKAR 249			
404 TTITSVLPRPALVASRVEYSKVLIL--	GGSGLSIGDAGEFDSSG-----	QAYKA--	451	
250 IRKIKLVETVEEML-----	SAIENELRSKKYDVVIMAAVSDFRPKIKAEKIKSGRS 303			
452 -----	MKEENVKTVLNNPNIASVQTNVEYGLQAD-----		480	
304 ITIELVAPXPKII--DIKELQRPVFLVGRKAEKSKKLEIEGKRQIERAKADLYVNTLE 362				
481 -IYFLPIPPQFTEVKEKQDPGLLIGMGQIALNCVGELEFRGLVKEGYKVLGTSTVE 539				
363 AFGESENOVVLIGRDTFKELPKKKRLEARI 394				
540 SIMATED-----	ROLFSKLNLEINKEI 561			

Thu Jan 31 13:32:15 2002

us-08-957-709-19_1.rsp

Page 10

Job time: 39 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:07:53 ; Search time 21.39 seconds
(without alignments)
1435.173 Million cell updates/sec

Title: US-08-957-709-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATKSRKLVGKK.....RMKKRELAEIWDIEIKXLS 403
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1667.5	83.5	401	2	pantothenate metab
2	1649.5	82.6	401	2	dna/pantothenate m
3	735	36.8	404	2	pantothenate metab
4	731	36.6	386	2	pantothenate metab
5	627.5	31.4	403	2	pantothenate metab
6	626	31.3	392	2	pantothenate metab
7	543.5	27.2	390	2	pantothenate metab
8	535.5	26.8	388	2	pantothenate metab
9	535	26.8	437	2	pantothenate metab
10	494.5	24.7	404	2	probable dna/panto
11	490.5	24.5	399	2	flavoprotein dfp
12	483.5	24.2	402	2	DNA/pantothenate m
13	482.5	24.1	394	2	DNA/pantothenate m
14	473.5	23.7	406	2	pantothenate metab
15	473	23.7	394	2	pantothenate metab
16	464	23.2	394	2	pantothenate metab
17	459.5	23.0	400	2	hypothetical prote
18	447	22.4	422	2	DNA/pantothenate m
19	439	22.0	402	2	pantothenate metab
20	430.5	21.5	418	2	pantothenate metab
21	424.5	21.2	430	1	pantothenate metab
22	424.5	21.2	430	2	hypothetical prote
23	404.5	20.2	429	2	DNA/pantothenate m
24	351	17.6	425	2	pantothenate metab
25	350	17.5	425	2	pantothenate metab
26	301	15.1	384	2	DNA/pantothenate
27	224.5	11.2	231	2	flavoprotein [impo
28	223	11.2	178	2	pantothenate metab
29	176.5	8.8	625	2	thymidylate syntha

ALIGNMENTS

RESULT 1

G71018

pantothenate metabolism flavoprotein dfp homolog PH1444 - Pyrococcus horikoshii
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: G71018

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Se M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: G71018

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-401 <KAW>

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30551.1; PID:g3257868

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH1444

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 83.5%; Score 1667.5; DB 2; Length 401;
Best Local Similarity 83.4%; Pred. No. 7e-94;
Matches 337; Conservative 33; Mismatches 29; Indels 5; Gaps 4;

Qy 1 MLHHVKLIYATKSRKLVGKKIVXXPGSIAALD-VKACEGLIRHGAEVHAVMSEAAATKII 59

Db 1 MLNHVKRIYAKSRKLVGKKIVLAIPGSAIVAEVCKLARELIRHGAEVHAVMTPSATKII 60

Qy 60 HPYAWNLTPGPNVTEITGFTIEHVELAGEHENKADLIIVCPATANTISKIACGIDDTPT 119

Db 61 HPYAMEFATGPNVTEITGFTIEHVELAGEHENKADLIIVCPATANTISKIACGIDDTPT 120

Qy 120 TVVTTAFPHIPIMAPAMHETMYRHPVIRENTERLKKLVGEFIEGRIGRAKVASIDEI 179

Db 121 TVVTTAFPHIPIMAPAMHESYKHPVIRENTERLKKLVGEFIEGRIGRAKVASIDEI 180

Qy 180 VYRVIKLHKHTLEGKRVLVTAGATREYIDPFRITFNASSGKMGVALAEADFRGA-VTL 238

Db 181 VYRVIRKLHPKTLGKRVLVTAGATREYIDPFRITFNASSGKMGVALAEAEFRGA-VTL 240

Qy 239 IRTKGSVKAFRIRKTLKVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKKAEKGI 298

Db 241 IRTKGSVSNFVENQI--QVETVEEMLAAIEKELTEKKYDVVIMAAVSDFKPKIAEKDI 298

Qy 299 KSGRSITIELVXPXNKIIDRIKEIQPNVFLYGFKAETSKELIEEGKQIERAKADLVVG 358

Db 299 KSNKSITIELVP-NPKIIDRIKEIQPDVFLYGFKAETTMEXLIQEAKKOIERAKSDIVIG 357

hypothetical prote
lantibiotic epider
hypothetical prote
hal3 protein - yea
SIS2 protein - yea
hypothetical prote
carbamoyl-phosphat
hypothetical prote
translation initia
chromosome assembl
probable membrane
conserved hypother
hypothetical prote
carbamoyl-phosphat
carbamoyl-phosphat
conserved hypother

C;Accession: C70201
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kervage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
A;Reference number: A70100; MUID:98065943
A;Accession: C70201
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-390 <KAE>
A;Cross-references: GB:AF001179; GB:AE000783; NID:92688738; PIDN:AAC67145.1; PID:g268873
A;Experimental source: strain B31
C;Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 27.2%; Score 543.5; DB 2; Length 390;
Best Local Similarity 36.3%; Pred. No. 7.3e-26;
Matches 144; Conservative 81; Mismatches 135; Indels 37; Gaps 14;
QY 19 KKIYXXPGSIAAL-DVKACEGLIRHGAHVAVHSEAAATKIIHPYAWNLPTGNFVITEIT 77
Db 5 KHLIGIGGGIYKSVYVSSLVKLVKVKVITQNTAKFTPLTLETISKNNIITNLW 64
QY 78 GF----IEHVELAGEHENKADLILVCPATANTISKIAGIDDTPTVTTVTTAFPHIPMI 133
Db 65 DLHDNEVEHikia---KWAHLILVIPATYNTISKIAGIADALTIIISAS--TAPTYF 118
QY 134 APAMHETMYRPIVRENIERIKLGVFIPRIEFG-----RAKVASIDEIVYRVIK 185
Db 119 AIAMNNIMYNPILKENIKKLTNYKFIET--DKGFLACSSNALGRKLNEDRIKIILN 176
QY 186 KLHKT-LEGRVLVTAGATREYIDPRTFTNASSGKMGVALAEADPRGA-VTLI---R 240
Db 177 EFNQYLNKAKILITSTELIDPRTFTNASSGKMGVALAEADPRGA-VTLI---R 236
QY 241 TKGSVKAPRIKIKLVETEMLSATENELRSKYDVVIMAAVSPRPKIKAEGLIKS 300
Db 237 NENDEGVNIIRIKITAMDMYKALYI-----NKFEIIGAAVADPKPHIFNSKIK 290
QY 301 GR--SITELVXPXPKIIDRIKETO-PNVFLVGPKAETSEKLEIEGKRQIERAKDLV 357
Db 291 KNINLYIKLV-KNPDIIOHGNKLNQIVIGFCAENSK-NLIQAKERIKLKNLDFII 348
QY 358 GNTLEAFGESENVVLIGRDTFKELPKMKKRELAERI 394
Db 349 ANELYFGSKLNKYIINKQSIKELPEMERSEVAKEI 385

RESULT 8
pantothenate metabolism flavoprotein dfp - Aquifex aeolicus
N;Alternate names: probable aspartate l-decarboxylase activase
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 29-Sep-1999
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
A;Reference number: A70300; MUID:98196666
A;Accession: B70371
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-388 <KQF>
A;Cross-references: GB:AE000708; NID:92983356; PIDN:AAC06944.1; PID:g2983357; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene dfp
C;Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 26.8%; Score 535.5; DB 2; Length 388;
Best Local Similarity 37.2%; Pred. No. 2.2e-25;
Matches 142; Conservative 74; Mismatches 137; Indels 29; Gaps 13;
QY 27 GSIAALDVKACE---GLIRHGAHVAVHSEAAATKIIHPYAWNLPTGNFVITEITGFIE-- 81
Db 9 GGIIAS--YKVGELVRELKRKGHSVKTILTPFAEKFMSPLTFTQISGNKAYTDKOWEEPL 66
QY 82 -HVELAGEHENKADLILVCPATANTISKIAGIDDTPTVTTVTTAFPHIPIMIPAMHET 140
Db 67 AHINLA---RWADVLIAPATANTIAKIANGISDNLLTTTI-LAYGR-PULVAPAMTV 120
QY 141 MYRPIVRENIERIKLGVFIPRI-----EGRKAVASIDEIVYRVIKLHKHTLEG 194
Db 121 MYKSPSTQENLKLKEWHVIEPEFVGLACKVEGEGKLASIDRLIDWIYVSEKPLKG 180
QY 195 KRVLVTAGATREYIDPRTFTNASSGKMGVALAEADPRGAVTLIRTKGSKVAKRIRIK 254
Db 181 KKVLTICGATKEFDIPVFISYSSGEMGFSLARIFRWKGAEVK-ITAGTTTAKPEPVE 239
QY 255 -LKVETVEEMLSATENELRSKY-DVVTMAAAVSDFRPKIKAEGLIKSGRSITIELVFXN 312
Db 240 IIRVQTTTEMR--EKLVEHYWDIVVMAAAVADFPKSVKSEKIKRDKLLELV-KN 295
QY 313 PKIIDRIKETOPNVFLVGPKAETSEKLEIEGKRQIERAKDLVGVNTLEAFGESENOV 372
Db 296 PDILEELGRKKYKILVGFALLES--DNLEIYAREKLEKRLNLDLIVANPVKVGAKREHGY 353
QY 373 LIGRDTFKELPKMKKRELAERI 394
Db 354 LITKDOIQVELPKGSKLENAREFI 375

RESULT 9
probable DNA/pantothenate metabolism flavoprotein APE1959 - Aeropyrum pernix (strain
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: A72498
R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aero*
A;Reference number: A72450; MUID:99310339
A;Accession: A72498
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <KAW>
A;Cross-references: DDBJ:AP000063; NID:95105654; PIDN:BAA80969.1; PID:g5105657
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1959
C;Superfamily: pantothenate metabolism flavoprotein dfp
Query Match 26.8%; Score 535; DB 2; Length 437;
Best Local Similarity 36.8%; Pred. No. 2.7e-25;
Matches 150; Conservative 66; Mismatches 152; Indels 40; Gaps 14;
QY 18 GKKIVXXPGSIA---ALDVKACEGLIRHGAHVAVHSEAAATKIIHPYAWNLPTGNFVIT 74
Db 47 KTKILVGTATVALYRSIDL--AWLLRGARVITVMPPEAKLVSPMFHWASGGPVYT 104
QY 75 EITGFIEHVELAGEHENKADLILVCPATANTISKIAGIDDTPTVTTVTTAFPH-IPIMI 133
Db 105 GFTGGVHEISIA---RAASVAVAPATLSTLAKIAGVVDNPVALAASVIMGYGKRVIA 160
QY 134 APAMHETMYRPIVRENIERIKLGVFIPRIEGRKAVASIDEIVYRVIKLHK--KT 191
Db 161 VPAMHGNMYESPOAREVVDRLRSGVLVDPKIEGGVAKYPTDHAVGRITAAQAKGLRD 220
QY 192 LEGRVLVTAGATREYIDPRTFTNASSGKMGVALAEADPRGA-VTLIRTKGSKVAKAFRI 250

Db 221 LEGIRALVTLGSTREWDIDRVFISNPSSGVMLGAAALEYARGAEVDVAGYTSVEIPLH 280
Qy 251 RKIKLVETVEEMLSAIENELRSKYDVVIMAAVSDFRPKIKAEGLKSGRSITIELVVP 310
Db 281 FN-TYKTEEDMAAAVELTSKREYDAVAAAPVDFRPAAGFGRKSGRLVLELP 339
Qy 311 XNPKIIRIKIOPNVFLVGFKAEE--TSKEKLIIEGKQRIERAKADLVVGNLTLEAFG--- 365
Db 340 -TPKVLEGIAR-RPKV-LVAFAAEVDNLDLSRPALEKMEKYDADLVVANRVGVEGVGF 396
Qy 366 -----SEENQVVLIGRDFTEKLPKMKKRELAERIWDEIEKXLS 403
Db 397 ASPLLDVLMLDKSGEAVLKG-SFHKEI-----VAAVIADIEIAKLKS 436

RESULT 10
F83963
flavoprotein dfp [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000
C:Accession: F83963
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314
A:Accession: F83963
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BA06229.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: dfp
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 24.7% Score 494.5; DB 2; Length 404;
Best Local Similarity 33.5%; Pred. No. 7e-23;
Matches 138; Conservative 84; Mismatches 149; Indels 41; Gaps 15;

Qy 16 LVGKKIYXXPGSIAALDVKA-CEGLRHGAHVHVMSEAATKIIHPYAWNLPNGPVIT 74
Db 2 LOGKRVVLGVSGGIAAFKSAFASKLVQAGAEVVMTEGAKKFVTLTFOALTRHPVYD 61
Qy 75 EI-----TGFEHVELAGEHENKADILVCPATANTISKIACGIDDTPTVTWTTAPPHI 129
Db 62 DTFSEPPSEIAHLQLA-----DWADVIIIAATANLICKLANGVADMLSTMLLAT--KA 115
Qy 130 PIMIAPAMHETMYRHPVIVRENIERLKLGVFEFIGPRIEE-----GRAKVASIDEIVRV 183
Db 116 PIYLAPAMNVNMYEHPAVORNNQQLAKDGYRLLEPGAGYLACGWIGKGRMPEPEDLLKTI 175
Qy 184 IKKLH-----KKTLEGRVLVTAGATREYIDPIREITNASSGKMGVALAEAA-DFRGAVTL 238
Db 176 --EVHFTPPSSLAGKKIIVITAGPTQETIDPIREFTNRSKMGKGYALTAKAARDFGGNVTL 233
Qy 239 IRTKGSV-KAFRIKIKLVETVEEMLSAIENELRSKYDVVIMAAVSDFRPKIKAEGLK 297
Db 234 ISGPTSLEKPDGVSVNVK--SAQDMYEAFLAEFSDA--DVVIKTAADVADYRPRVHQQ 289
Qy 298 IKSGRSITIELVPXNPKIIRIKIOPNVFLVGFKAETSKEKLEEGKRQIERAKADLVV 357
Db 290 VKKGDWGVIELERTVDILKTLGEEKESQFLVGFAAES--QEVETYAQKKLKEKNADIV 347
Qy 358 GNTL-----EAFGSEENQV-VLIGRDFTEKLPKMKKRELAER-----IWDEIEKX 400
Db 348 ANNVTEGAGFQDTNRTVTVYFGQGVKPLPLMTKDEVAHRLIMMISEQLEK 399

RESULT 11
A82351
DNA/pantothenate metabolism flavoprotein VC0215 [imported] - Vibrio cholerae (strain N16
C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: A82351
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: A82351
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-399 <HEI>
A:Cross-references: GB:AE004111; GB:AE003852; NID:g9654614; PIDN:AAF93391.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0215
A:Map position: 1
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 24.5% Score 490.5; DB 2; Length 399;
Best Local Similarity 35.18; Pred. No. 1.2e-22;
Matches 145; Conservative 69; Mismatches 160; Indels 39; Gaps 15;

Qy 14 RLVGKKIYXXPGSIAALD-VKACEGLRHGAHVHVMSEAATKIIHPYAWNLPNGPV 72
Db 2 QSLAGKILLGLISGIAAYKCAELTRRLVERGATVQVVMTHAAKEFITPLTMQAVSRPV 61
Qy 73 ITEI-----TGFEHVELAGEHENKADILVCPATANTISKIACGIDDTPTVTWTT 125
Db 62 SDSLLDPAAEASMGHIELA----KWADLVLLAPATADLIARMAAGMNDLTTILATSA 117
Qy 126 FPHIPIMIAPAMHETMYRHPVIVRENIERLKLGVFEFIGPRIEE-----GRAKVASIDEI 179
Db 118 ----PVAIAPAMNQOMYRNIAQTQENLQILIRRGILTWGPAAGCAQCGVGRMLPELMEL 173
Qy 180 VYRVIKLLHKKLTLEGKRVLTAGATREYIDPIREITNASSGKMGVALAEAAEDFRGA-VTL 238
Db 174 VAHCENFPAPKILVKGKRVLTAGPTREALDPVRYITNHSKMGKGFALAKAAQGLADVTL 233
Qy 239 IRTKGSV-KAFRIKIKLVETVEEMLSAIENELRSKYDVVIMAAVSDFRPKIKAEGLK 297
Db 234 V--SGPVHLPTPGVGNRIDVQSGLEMHSAVMKEATS--HQIFIACAAVADYRPTVASQK 289
Qy 298 IKSGR--SITIELVPXNPKIIRIKIOPNV-VFLVGFKAETSKEKLEEGKRQIERAKA 353
Db 290 IKSRDNDTLIEWV-KNPDIVASVAALTENRPTVGFPAETQDVEYARSK--LYRKNL 346
Qy 354 DLVVGNTL-----EAFGSEENQVVLIGRDFTEKLPKMKKRELAERIWDEIEKX 402
Db 347 DMICANDVSIAGQGFNSNDNALTFLWKEGQHSPLITSKDALASAVMHLIHEOM 399

RESULT 12
B82982
DNA/pantothenate metabolism flavoprotein PA5320 [imported] - Pseudomonas aeruginosa (C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B82982
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: B82982
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-402 <STO>
A:Cross-references: GB:AE004944; GB:AE004091; NID:g9951628; PIDN:AG08705.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: dfp; PA5320

C:Superfamily: pantothenate metabolism flavoprotein dff

Query Match 24.2%; Score 483.5; DB 2; Length 402;
 Best Local Similarity 33.4%; Pred. No. 3.2e-22;
 Matches 140; Conservative 81; Mismatches 147; Indels 51; Gaps 17;

QY 14 RKLKGVKIVXXPGSTALDVKACGLIR----HGAEVHAVNSEAATKIIHPYANWLPTG 69
 DB 2 ORLYKRIRLVVGGGGIAA--YKSAB-LVRLRDQGAERVVVMTQGRFITPLTLQALSG 58
 QY 70 NPVITEI-----TGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVT 124
 DB 59 HPVHTDLDPAAEAAMGHIELA----RWADLVLIAPATADLMAVQGVANDLLTLVLA 114
 QY 125 APFHPIMIAFAMHETMYRHPVRENIERLKLGVFEFTGPRI-----EEGRAKVASIDE 178
 DB 115 T--DAQIALAPAMQAWRDATQANAEALLRQRFHLEFGPAAGSQACGDVGLGRMLEAEE 172
 QY 179 IVRVYIKLHLKRTLEGKRVLTAGATREYIDPRTITNASSGKMGVALAEADFRGA-VT 237
 DB 173 LAQRAADCFOQALTGHVHLITAGPTQENIDPVRVITNHSSGKMGFALAEAAVEAGARVT 232
 QY 238 LIRTKGSV---KAFRIKIKLKVETVEEMLSAIENELSKKYDVVIMAAVSDFRPKIKA 294
 DB 233 LV--TGPVHLPTDVRQVR--DVVSARDMLAAE---AAMPDDLIIASAADVADYRPEVVA 285
 QY 295 EGKIK----SGRSITIELVPXPKIIDIKEIOPNVFLVGFKAETSKKLEEGKROIER 350
 DB 286 AHKLKDKPTSGELLQLV-LRNPDLATLAQREDRPFVSGFAAEI--ENLLDYAARKLKD 342
 QY 351 AKADLVGNTLE---ARGSENOVVLGRD---FTKELPKMKKRLAERIWIDEIK 400
 DB 343 KMLDLIVANDVANPSIGNSDENATVTDRLDHPSTFAQTSGKIARQLVAFIADRLNQ 401

RESULT 13
 A72223
 pantothenate metabolism flavoprotein dff homolog TM1687 - Thermotoga maritima (strain MS
 N:Alternate names: probable aspartate 1-decarboxylase activase
 C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: A72223
 R:Neilson, K.E.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 Garrett, M.M.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; MUID:99287316
 A:Accession: A72223
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-394 <ARN>
 A:Cross-references: GB:AE001809; GB:AE000512; NID:g4982257; PIDN:AAD36754.1; PID:g498226
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM1687
 C:Superfamily: pantothenate metabolism flavoprotein dff

Query Match 24.1%; Score 482.5; DB 2; Length 394;
 Best Local Similarity 36.2%; Pred. No. 3.6e-22;
 Matches 144; Conservative 67; Mismatches 154; Indels 33; Gaps 14;

QY 20 KIVXXPGSIA---ALDVKACGLIRHGAEVHAVNSEAATKIIHPYANWLPTGNPITE- 75
 DB 2 RIVLVSSGGIAIYKAVDL--ASKLRKEGHELHVVTPTDTRMISPVFVS-AVGNCVSVDH 58
 QY 76 ----ITGFIHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTAFPHIPI 131
 DB 59 WMDVRNGWIPHEIS----RTADLVVAPATANTISKIANGIADN-LJTLVALAFDKDAK 113
 QY 132 MIAPAMHETMYRHPVRENIERLKLGVFEFTGPRIEEGRAKVASIDEIVY----RVIKKL 187

Db 114 ILVPTMTNMYSNKLFQENLEKKNNGFVVEP--EGHLAGCEVKGGRYPENEKIVEAV 171
 QY 188 H----KYLEGKRVLTAGATREYIDPRTITNASSGKMGVALAEADFRGA-VTLIRTK 242
 DB 172 HLLTTPFKLAGKKVLITAGPTREIDAVRFTNASSGKMGYALATVAKRMGARVSLVSGP 231
 QY 243 GSVKAFRIKIKLKVETVEEMLSAIENELSKKYDVVIMAAVSDFRPKIKAEKIKSGR 302
 DB 232 TSLKPPYVDFEVKVESAEEMVEYMK--RFEDTDIVIMNAVAGDYRPKKVEGKLKTE 289
 QY 303 STTIELVPXPKIIDIKEIOPNVFLVGFKAETSKKLEEGKROIERAKADLVVGN-TL 361
 DB 290 NDVLVHLERTDKILGORKSNQILVGFAAEV--EENEENAVKKRLKRLNLLDLVNDAR 347
 QY 362 EAFGSENOVVLGRD--FTKELPKMKKRLAERIWIDEI 398
 DB 348 KAFSSDRVEVYIGRDFIKRIDEDKIRVAGGILDV 385

RESULT 14
 D69878
 pantothenate metabolism flavoprotein dff homolog yloI - Bacillus subtilis
 N:Alternate names: probable aspartate 1-decarboxylase activase
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: D69878
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emmergon, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y, M.; Ogawa, K.; Ogiwara, C.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivoita, C.; Roche, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Se
 kuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winkler, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033
 A:Accession: D69878
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-406 <GUN>
 A:Cross-references: GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAB13443.1; PID:g26339
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yloI
 C:Superfamily: pantothenate metabolism flavoprotein dff

Query Match 23.7%; Score 473.5; DB 2; Length 406;
 Best Local Similarity 34.1%; Pred. No. 1.3e-21;
 Matches 143; Conservative 73; Mismatches 142; Indels 61; Gaps 17;

QY 16 LVGKKIVXXPGSTALDVKAC---EGLIRHGAEVHAVNSEAATKIIHPYANWLPTGNPV 72
 DB 2 LNNRVNLCVSGGIIV--YKACALTSKLVQAGANVKVIMTESACRFVSLTFOALSREHV 59
 QY 73 ITEI-----TCFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVTVTVTAPP 127
 DB 60 YDTDFKEQNPSVISHIDAA----DWADLIIVAPATANVIGLANCIADMLTTLTAA-- 113
 QY 128 HIPMIAFAMHETMYRHPVRENIERLKLGVFEFTGPRIEE-----GRAKVASIDEI 179
 DB 114 TAPVVIAPAMVHMVHDHPAVKRNISLVLYODGYCFLEP--SEGYLACGVGKGRLEPENI 171
 QY 180 VYRVIKLHKKT---LEGRKRVLTAGATREYIDPRTITNASSGKMGVALAEADFRGA- 235
 DB 172 VKLAEKHEFAETSAPLEGHVVITAGPTREIDPVRFTNKSSTGKMGYALAEAAVQLGAR 231

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:13 ; Search time 46.78 seconds
(without alignments)
122.268 Million cell updates/sec

Title: US-08-957-709-71

Perfect score: 156

Sequence: 1 MLLPDWKIRKEILIEPFSEE.....PYRGNYGSTRLEAFSKRKL 156

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37	23.7	156	1 DCD_PYRHO	O57706 pyrococcus
2	22	14.1	154	1 DCD_PYRAB	Q9uxs8 pyrococcus
3	8	5.1	420	1 ODO2_BUCAI	P57389 buchnera ap
4	7	4.5	133	1 FLIS_BACSU	P39739 bacillus su
5	7	4.5	178	1 AIL_YEREN	P16454 versinia en
6	7	4.5	182	1 AIL_YERPS	Q56957 versinia ps
7	7	4.5	215	1 TDH_SULSO	P95895 sulfolobus
8	7	4.5	237	1 HIS4_METJA	Q58927 methanococc
9	7	4.5	279	1 NAPG_HAEIN	P44652 haemophilus
10	7	4.5	327	1 CYSM_AQUAE	O67507 aquifex aeo
11	7	4.5	439	1 QUIX_ACTICA	Q43923 acinetobact
12	7	4.5	460	1 ALN_YEAST	P32375 saccharomyc
13	7	4.5	682	1 NTP2_FOPPV	O72904 fowlpox vir
14	7	4.5	684	1 NTP2_MCV1	Q98218 molluscum c
15	7	4.5	913	1 UBPK_HUMAN	Q9Y2K6 homo sapien
16	7	4.5	1113	1 Y14O_MYCGE	P47386 mycoplasma
17	7	4.5	1121	1 DPOL_ADEG1	O64751 avian adeno
18	7	4.5	1217	1 SYV_FUGRU	P49696 fugu rubrip
19	7	4.5	3210	1 CENF_HUMAN	P49454 homo sapien
20	6	3.8	37	1 RK36_EUGGR	P21532 euglena gra
21	6	3.8	62	1 RS21_ANAVA	P49224 anabaena va
22	6	3.8	85	1 HKL8_MAIZE	P56666 zea mays (m
23	6	3.8	94	1 EM2_WHEAT	P22701 triticum ae
24	6	3.8	110	1 YK21_CAEEL	P34328 caenorhabdi
25	6	3.8	114	1 NU3M_XENLA	P03900 xenopus lae
26	6	3.8	119	1 RL24_SULSO	Q9ux95 sulfolobus
27	6	3.8	124	1 VATF_MANSE	P31478 manduca sex
28	6	3.8	136	1 VG46_BPMD2	O64237 mycobacteri
29	6	3.8	171	1 OMPX_ECOLI	P36546 escherichia
30	6	3.8	172	1 OMPX_ENTCL	P25253 enterobacte
31	6	3.8	174	1 YRKL_BACSU	P54439 bacillus su
32	6	3.8	175	1 RPI8_RAT	P71940 rattus norv
33	6	3.8	183	1 DCD_THEAC	Q9hkk0 thermoplasm

ALIGNMENTS

RESULT 1

ID	DCD_PYRHO	STANDARD;	PRT;	156 AA.
AC	O57706;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE).			
GN	DCD OR PH1997.			
OS	Pyrococcus horikoshii.			
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.			
OX	NCBI_TaxID=53953;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN=OT3;			
RX	MEDLINE=983444137; PubMed=9679194;			
RA	Kawarayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.			
RA	"Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3."			
RL	DNA Res. 5:55-76(1998).			
CC	-I- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).			
CC	-I- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AP000007; BAA31124.1; -			
DR	InterPro; IPR003232; dCTP_deaminase.			
DR	InterPro; IPR001428; dUTPase.			
DR	Pfam; PF00692; dUTPase; 1.			
DR	ProDom; PD004900; dCTP_deaminase; 1.			
KW	Hydrolase; Complete proteome.			
SQ	SEQUENCE 156 AA; 17871 MW; 96B2C2C50393A985 CRC64;			

Query Match 23.7%; Score 37; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 3.9e-31;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIEPFSEESLOPAGYDLRVGREAFV 37

Db 1 MLLPDWKIRKEILIEPFSEESLOPAGYDLRVGREAFV 37

RESULT 2

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; Z31376; CAAB3249.1; -
DR EMBL; U56901; AAC44954.1; -
DR EMBL; Z99122; CAB15550.1; -
DR Subtilist; BG10922; flis.
DR InterPro; IPR003713; flis.
DR Pfam; PF02561; flis; 1.
KW Flagella; Complete proteome.
SQ SEQUENCE 133 AA; 15131 MW; E0E80476A96F14D3 CRC64;

Query Match 4.5%; Score 7; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 LTLMLYN 108
Db 23 LTLMLYN 29
|||||||

RESULT 5
AIL_YEREN STANDARD; PRT; 178 AA.
AC F16434;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATTACHMENT INVASION LOCUS PROTEIN PRECURSOR.
GN AIL.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
RX MEDLINE-90130261; PubMed-1688838;
RA Miller V.L., Bliska J.B., Falkow S.;
RT "Nucleotide sequence of the Yersinia enterocolitica ail gene and
RT characterization of the Ail protein product.";
RL J. Bacteriol. 172:1062-1069(1990).
CC -!- FUNCTION: THIS MEMBRANE-ASSOCIATED PROTEIN PROMOTES INVASION
CC OF PATHOGENIC BACTERIA INTO EUKARYOTIC CELLS BY AN UNKNOWN
CC MECHANISM.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOM FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M29945; AAA88694.1; -
DR PIR; A35123; A35123.
DR InterPro; IPR000758; Enterovir_OMP.
DR PRINTS; PR00316; ENTEROVIR_OMP.
DR PROSITE; PS00694; ENT_VIR_OMP_1; 1.
DR PROSITE; PS00695; ENT_VIR_OMP_2; 1.
KW Outer membrane; Transmembrane; Signal; Virulence.
FT SIGNAL 1 23
FT CHAIN 24 178 ATTACHMENT INVASION LOCUS PROTEIN.
SQ SEQUENCE 178 AA; 19548 MW; D46FC92E2524F162 CRC64;

Query Match 4.5%; Score 7; DB 1; Length 178;

Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92
Db 65 GVIGSFA 71
|||||||

RESULT 6
AIL_YERPS STANDARD; PRT; 182 AA.
AC Q56957;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATTACHMENT INVASION LOCUS PROTEIN PRECURSOR.
GN AIL.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YPIII;
RX MEDLINE-96294755; PubMed-8698470;
RA Yang Y., Merriam J.J., Mueller J.P., Isberg R.R.;
RT "The psa locus is responsible for thermoinducible binding of Yersinia
RT pseudotuberculosis to cultured cells.";
RL Infect. Immun. 64:2483-2489(1996).
CC -!- FUNCTION: THIS MEMBRANE-ASSOCIATED PROTEIN PROMOTES INVASION
CC OF PATHOGENIC BACTERIA INTO EUKARYOTIC CELLS BY AN UNKNOWN
CC MECHANISM (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOM FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; L49439; AAB36601.1; -
DR InterPro; IPR000758; Enterovir_OMP.
DR PRINTS; PR00316; ENTEROVIR_OMP.
DR PROSITE; PS00694; ENT_VIR_OMP_1; 1.
DR PROSITE; PS00695; ENT_VIR_OMP_2; 1.
KW Outer membrane; Transmembrane; Signal; Virulence.
FT SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 182 ATTACHMENT INVASION LOCUS PROTEIN.
SQ SEQUENCE 182 AA; 20192 MW; EE7AAA20E4D8975C CRC64;

Query Match 4.5%; Score 7; DB 1; Length 182;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92
Db 68 GVIGSFA 74
|||||||

RESULT 7
TDXH_SULSO STANDARD; PRT; 215 AA.
AC P95895;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PEROXIREDOXIN.
GN SSO2121 OR C02015 OR C02_016.
OS Sulfolobus solfataricus.

```

OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RC MEDLINE=97055432; PubMed=8899719;
RA Sensen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
RA Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
RA Doolittle W.F., Ragan M.A., Charlebois R.L.;
RT "Organizational characteristics and information content of an
RT archaeal genome: 156 kb of sequence from Sulfolobus solfataricus
RT P2.";
RL Mol. Microbiol. 22:175-191(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RC MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Weiler C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -1- SIMILARITY: BELONGS TO THE AHPC-TSA FAMILY. TDXH SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y08256; CAA69447.1; -
CC EMBL: AL512975; CAC23765.1; -
CC EMBL: AE006819; AAK42300.1; -
CC HSP: P30041; IPRX.
CC InterPro: IPR00866; Ahpc-TSA.
CC Pfam: PF00578; Ahpc-TSA; 1.
CC Antioxidant; Complete proteome.
FT ACT_SITE 49 49 BY SIMILARITY.
SQ SEQUENCE 215 AA: 24745 MW; 5210477EF4C9607B CRC64;

Query Match 4.5%; Score 7; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIKLPPD 71
DB 24 RIKLPPD 30

RESULT 8
ID HIS4_METJA STANDARD; PRT; 237 AA.
AC Q58927;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHORIBOSYLFORMIMINO-5-AMINOIMIDAZOLE CARBOXAMIDE RIBOTIDE
DE ISOMERASE (EC 5.3.1.16).
GN HIS4 OR M11532
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

```

```

RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: N-(5'-PHOSPHO-D-RIBOSYLFORMIMINO)-5-AMINO-1'-
CC (5'-PHOSPHORIBOSYL)-4-IMIDAZOLECARBOXAMIDE - N-(5'-PHOSPHO-D-1'-
CC RIBULOSYLFORMIMINO)-5-AMINO-1-(5'-PHOSPHORIBOSYL)-4-
CC IMIDAZOLECARBOXAMIDE.
CC -1- PATHWAY: FOURTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67594; AAB99553.1; -
CC TIGR: MJ1532; -
CC InterPro: IPR003009; FMN enzyme.
CC InterPro: IPR000570; His-biosynth.
CC Pfam: PF00977; His-biosynth; 1.
CC Isomerase; Histidine biosynthesis; Complete proteome.
SQ SEQUENCE 237 AA: 26132 MW; D1EE3AD24C0AC2A CRC64;

Query Match 4.5%; Score 7; DB 1; Length 237;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 KEGKVI 52
DB 131 KEGKVI 137

RESULT 9
ID NAPG_HAEIN STANDARD; PRT; 279 AA.
AC P44652;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FERREDOXIN-TYPE PROTEIN NAPG HOMOLOG.
GN NAPG OR HI0345.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus

```

```

RT RL Science 269:496-512(1995).
CC -|- FUNCTION: INVOLVED IN ELECTRON TRANSFER.
CC -|- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32719; AAC22006.1; -.
DR HSPG; P00198; 1FCA.
DR TIGR; H0345; -.
DR InterPro: IPR001450; 4FE4S_ferredoxin.
DR PRAM; PF00037; fer4; 2.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW Electron transport; Iron-sulfur; 4Fe-4S; Complete proteome.
FT METAL 72 72 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 75 75 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 78 78 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 82 82 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT METAL 110 110 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 113 113 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 118 118 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 122 122 IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT METAL 150 150 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 158 158 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 161 161 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 165 165 IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT METAL 197 197 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 200 200 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 203 203 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT METAL 207 207 IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 279 AA; 30227 MW; F77B6801E220955A CRC64;

Query Match 4.5%; Score 7; DB 1; Length 279;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 SLAREGV 87
| | | | |
DB 48 SLAREGV 54

RESULT 10
CYSM_AQUAE STANDARD; PRT; 327 AA.
AC O67507;
DT 30-MAY-2000 (Rel. 39, Created)
DT 20-AUG-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYSTEINE SYNTHASE (EC 4.2.99.8) (O-ACETYL-SERINE SULFHYDRYLASE)
DE (O-ACETYL-SERINE (THIOL)-LYASE) (CSASE).
GN CYSM OR AQ_1556.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus."
RL Nature 392:353-358(1998).
CC -|- CATALYTIC ACTIVITY: O-ACETYL-L-SERINE + H(2)S = L-CYSTEINE +

```

```

CC ACETATE.
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -|- PATHWAY: CYSTEINE BIOSYNTHESIS.
CC -|- SIMILARITY: BELONGS TO THE CYSTEINE SYNTHASE/CYSTATHIONINE BETA-
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE00745; AAC07459.1; -.
DR InterPro: IPR001216; Cys_synthase.
DR InterPro: IPR001926; PALP.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00901; CYS_SYNTHASE; 1.
KW Lyase; Cysteine biosynthesis; Pyridoxal phosphate; Complete proteome.
FT BINDING 65 65 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 327 AA; 36155 MW; 54FC4338949FB40E CRC64;

Query Match 4.5%; Score 7; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 KEGKVI 52
| | | | |
DB 83 KEGKVI 89

RESULT 11
QUIX_ACICA STANDARD; PRT; 439 AA.
AC Q43923;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE PORIN QUIX PRECURSOR.
GN QUIX.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OX Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BD413 / ADP1;
RX MEDLINE=96011389; PubMed=7592351;
RA Elsemore D.A., Ornston L.N.;
RT "Unusual ancestry of dehydratases associated with quinate catabolism
in Acinetobacter calcoaceticus."
RL J. Bacteriol. 177:5971-5978(1995).
CC -|- FUNCTION: COULD POTENTIALLY BE INVOLVED IN THE TRANSPORT OF
CC QUINATE OR SHIKIMATE.
CC -|- SIMILARITY: STRONG, TO PSEUDOMONAS AERUGINOSA OPRB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05770; AAC37160.1; -.
DR Transprot; Transmembrane; Porin; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 439 PUTATIVE PORIN QUIX.
SQ SEQUENCE 439 AA; 49309 MW; B332A322F3130F79 CRC64;

```

```
Query Match 4.5%; Score 7; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 61 LTLERIK 67
Db 110 LTLERIK 116

RESULT 12
ID ALN_YEAST STANDARD; PRT; 460 AA.
AC P32375;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALLANTOINASE (EC 3.5.2.5).
GN DAL1 OR YR027C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A. PubMed=1803816;
RX BUCKHOLZ R.G., Cooper T.G.;
RY "The allantoinase (DAL1) gene of Saccharomyces cerevisiae.";
RL Yeast 7:913-923(1991).
RN [2]
RN ERRATUM.
RP BUCKHOLZ R.G., Cooper T.G.;
RL Yeast 8:239-239(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Lewis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: UTILIZATION OF PURINES AS SECONDARY NITROGEN SOURCES.
CC -1- WHEN PRIMARY SOURCES ARE LIMITING.
CC -1- CATALYTIC ACTIVITY: ALLANTOIN + H(2)O = ALLANTOATE.
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- PATHWAY: FIRST STEP IN THE DEGRADATION OF ALLANTOIN (PURINE
CC CATABOLISM).
CC -1- SIMILARITY: BELONGS TO THE DHOASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M69294; AAA34553.1; -
CC PIR: 238061; CAA86187.1; -
CC PIR: S28649; S28649.
CC SCD: S48489; S48489.
CC SCD: S0001466; DAL1.
CC InterPro: IPR002195; Dihydroorotase.
CC Pfam: PF00744; Dihydroorotase_1.
CC PROSITE: PS00482; DIHYDROOROTASE_1; 1.
CC PROSITE: PS00483; DIHYDROOROTASE_2; FALSE_NEG.
CC Hydrolase; Purine metabolism; Zinc.
CC METAL 70
CC ZINC (POTENTIAL).
CC METAL 72
CC CONFLICT 438 460 VVYTNANGSVKTPLGQTLDSRR -> WYTRMPTESRKHWH
CC SEQUENCE 460 AA; 50126 MW; 887480809B6A8F42 CRC64;

Query Match 4.5%; Score 7; DB 1; Length 460;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 84 REGVIGS 90
Db 313 REGVIGS 319

RESULT 13
ID NTP2_FOWPV STANDARD; PRT; 682 AA.
AC O72904;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEOSIDE TRIPHOSPHATASE II (EC 3.6.1.15) (NUCLEOSIDE
DE TRIPHOSPHATE PHOSPHOHYDROLASE II) (NPH II).
GN F18R OR FV082.
OS Fowlpox virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HP-440;
RX MEDLINE=92410746; PubMed=1326827;
RA Binns M.M., Boursnell M.E.G., Skinner M.A.;
RT "Gene translocations in poxviruses: the fowlpox virus thymidine kinase
RT gene is flanked by 15 bp direct repeats and occupies the locus which
RT in vaccinia virus is occupied by the ribonucleotide reductase large
RT subunit gene.";
RL Virus Res. 24:161-172(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=20193820; PubMed=10729156;
RX Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RA "The genome of fowlpox virus.";
RL J. Virol. 74:3815-3831(2000).
CC -1- FUNCTION: ESSENTIAL FOR VIRAL REPLICATION. PLAYS AN IMPORTANT ROLE
CC DURING TRANSCRIPTION OF EARLY MRNAs, PRESUMABLY BY PREVENTING R-
CC LOOP FORMATION BEHIND THE ELONGATING RNA POLYMERASE. ACTS AS NTP-
CC DEPENDENT HELICASE THAT CATALYZES UNIDIRECTIONAL UNWINDING OF
CC 3' TAILED DUPLEX RNAs. MIGHT ALSO PLAY A ROLE IN THE EXPORT OF
CC NEWLY SYNTHESIZED MRNA CHAINS OUT OF THE CORE INTO THE CYTOPLASM.
CC REQUIRED FOR PROPAGATION OF VIRAL PARTICLES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + ORTHOPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAH
CC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: A7223385; CAA11299.1; -
CC EMBL: AF198100; AAF4426.1; -
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002484; DEAH_ATP_helcse.
CC InterPro: IPR001650; Helicase_C.
CC SMART: SM00490; HELIC_C; 1.
CC PROSITE: PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
CC ATP-binding; Helicase; Hydrolase; Transcription.
CC DOMAIN 281 516
CC NP_BIND 194 201 HELICASE.
CC SITE 303 306 DEXH BOX.
CC SEQUENCE 682 AA; 79807 MW; F394E69FD9745652 CRC64;
```

Query Match 4.5%; Score 7; DB 1; Length 682;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 PVELRYG 119
|||||
DB 263 PVELRYG 269

RESULT 14
NTP2_MCV1 STANDARD; PRT; 684 AA.
AC Q98218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE NUCLEOSIDE TRIPOSPHATASE II (EC 3.6.1.15) (NTPASE II) (NUCLEOSIDE
DE TRIPOSPHATE PHOSPHOHYDROLASE II) (NPH II).
GN 50R.
OS Molluscum contagiosum virus subtype 1 (MCV1).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Molluscipoxvirus.
OX NCBI_TaxID=10280;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=96325459; PubMed=8670425;
RX Senkevich T.G., Bugert J.J., Sisler J.R., Koonin E.V., Darai G.,
RA Moss B.;
RT "Genome sequence of a human tumorigenic poxvirus: prediction of
RT specific host response-evasion genes";
RL Science 273:813-816(1996).
CC -!- FUNCTION: ESSENTIAL FOR VIRAL REPLICATION. PLAYS AN IMPORTANT ROLE
CC DURING TRANSCRIPTION OF EARLY MRNAs. PRESUMABLY BY PREVENTING R-
CC LOOP FORMATION BEHIND THE ELONGATING RNA POLYMERASE. ACTS AS NTP-
CC DEPENDENT HELICASE THAT CATALYZES UNIDIRECTIONAL UNWINDING OF
CC 3' TAILED DUPLEX MRNAs. MIGHT ALSO PLAY A ROLE IN THE EXPORT OF
CC NEWLY SYNTHESIZED MRNA CHAINS OUT OF THE CORE INTO THE CYTOPLASM.
CC REQUIRED FOR PROPAGATION OF VIRAL PARTICLES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + ORTHOPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAD
CC SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: U60315; AAC55178.1;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002464; DEAD_ATP_helcse.
DR InterPro: IPR001650; Helicase_C.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; 1.
KW ATP-binding; Helicase; Hydrolase; Transcription.
FT DOMAIN 305 520 HELICASE.
FT NP_BIND 197 204 ATP (BY SIMILARITY).
FT SITE 308 311 DEXH BOX.
SQ SEQUENCE 684 AA; 76556 MW; 51072B193CCCT7284 CRC64;

Query Match 4.5%; Score 7; DB 1; Length 684;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 PVELRYG 119
|||||
DB 265 PVELRYG 271

RESULT 15
UBPK_HUMAN STANDARD; PRT; 913 AA.
ID AC Q9Y2K6; Q9UQN8; Q9UQPO;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 20 (EC 3.1.2.15) (UBIQUITIN
DE THIOLESTERASE 20) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 20)
DE (DEUBIQUITINATING ENZYME 20).
GN USP20 OR LSPR3A OR KIAA1003.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro";
RL DNA Res. 6:63-70(1999).
RN [2]
RN SEQUENCE OF 580-802 AND 838-913 FROM N.A.
RX MEDLINE=99299247; PubMed=10369878;
RA Gilley J., Fried M.;
RT "Extensive gene order differences within regions of conserved synten
RT between the Fugu and human genomes: implications for chromosomal
RT evolution and the cloning of disease genes";
RL Hum. Mol. Genet. 8:1313-1320(1999).
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
CC UBIQUITIN + A THIOLE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: AB023220; BAA76847.1;
DR EMBL; Y17457; CAB44350.1;
DR EMBL; Y17459; CAB44352.1;
DR MEROPS; C19.025;
DR InterPro: IPR001394; UCH-2.
DR InterPro: IPR001607; zf-UBP.
DR Pfam; PF00442; UCH-1; 1.
DR Pfam; PF00443; UCH-2; 1.
DR Pfam; PF02148; zf-UBP; 1.
DR SMART; SM00290; Znf-UBP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
KW Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family.
FT ACT_SITE 154 154 BY SIMILARITY.
FT ACT_SITE 634 634 BY SIMILARITY.
FT ACT_SITE 642 642 BY SIMILARITY.
FT CONFLICT 775 775 H -> Q (IN REF. 2).
FT CONFLICT 793 793 R -> M (IN REF. 2).
SQ SEQUENCE 913 AA; 101915 MW; 0FD618D51CFE99BD CRC64;

Query Match 4.5%; Score 7; DB 1; Length 913;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 EAFVKGK 40

Db 828 EAFVKGK 834
|||||||

Search completed: January 31, 2002, 13:39:15
Job time: 75 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:30 ; Search time 130.99 seconds
(without alignments)
450.017 Million cell updates/sec

Title: US-08-957-709-19
Perfect score: 403
Sequence: 1 MLHHVLIYATSRKLVGKK.....KMKRELARIWDEIEKXLS 403

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL_17.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp-organelle.*
 - 9: sp-phage.*
 - 10: sp-plant.*
 - 11: sp-rodent.*
 - 12: sp-virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*

* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	17.9	401	1	Q59114
2	60	14.9	401	1	Q9V0S3
3	12	3.0	394	2	Q9X213
4	10	2.5	394	2	Q9JVB7
5	10	2.5	394	2	Q9JTB7
6	10	2.5	399	2	Q9JTB7
7	9	2.2	382	1	Q9HH70
8	9	2.2	404	1	Q28628
9	8	2.0	127	5	Q9XUT2
10	8	2.0	142	11	Q9RIA1
11	8	2.0	169	10	Q9ZSW7
12	8	2.0	210	13	Q91596
13	8	2.0	227	2	Q9CKU9
14	8	2.0	355	11	P70441
15	8	2.0	396	2	Q9X8S5
16	8	2.0	481	2	Q25258
17	8	2.0	481	2	Q9ZLV5
18	8	2.0	481	2	Q9JMV2
19	8	2.0	498	10	Q9LJH1

20	8	2.0	794	2	Q9RIA3
21	8	2.0	1261	2	P73926
22	7	1.7	22	4	Q9BYX0
23	7	1.7	35	12	Q9YMA4
24	7	1.7	36	9	O80141
25	7	1.7	60	1	O26288
26	7	1.7	111	2	O66738
27	7	1.7	124	2	O06566
28	7	1.7	130	1	Q9UZZ9
29	7	1.7	138	11	Q9DA06
30	7	1.7	140	11	Q9D7H1
31	7	1.7	144	10	Q9SYF2
32	7	1.7	147	2	Q9CM27
33	7	1.7	188	2	O87200
34	7	1.7	194	4	O14878
35	7	1.7	194	4	Q9H3H8
36	7	1.7	194	4	Q9BY32
37	7	1.7	198	2	Q9F8S5
38	7	1.7	198	11	Q9D892
39	7	1.7	200	5	Q9V832
40	7	1.7	203	3	O14012
41	7	1.7	208	10	Q9FMY8
42	7	1.7	209	2	O31668
43	7	1.7	214	2	Q9JN89
44	7	1.7	220	1	Q9HK78
45	7	1.7	230	12	Q9IHR1

ALIGNMENTS

RESULT	1
O59114	PRELIMINARY; PRT; 401 AA.
ID	O59114
AC	O59114
DT	01-AUG-1998 (TREMBLrel. 07, Created)
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	401AA LONG HYPOTHETICAL PROTEIN DFP.
GN	PHI444.
OS	Pyrococcus horikoshii.
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX	NCBI_TaxID=53953;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=OT3;
RX	MEDLINE=98344137; PubMed=9679194;
RA	Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.F., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.;
RT	*Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.*;
RL	DNA Res. 5:55-76(1998).
DR	EMBL; AP000006; BAA30551.1;
DR	InterPro; IPR003382; Flavoprotein.
DR	Pfam; PF02441; Flavoprotein; 1.
KW	Complete proteome.
SQ	SEQUENCE 401 AA; 44624 MW; 26CBC5F523AE02C8 CRC64;

Query Match	17.9%; Score 72; DB 1; Length 401;
Best Local Similarity	100.0%; Pred. No. 3.1e-63;
Matches	72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	68 TGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACGIDTPTVTVTAPP 127
Db	69 TGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACGIDTPTVTVTAPP 128
Qy	128 HIPIMIAFAMHE 139

```

Db 129 HIPIMIAPAMHE 140
RESULT 2
Q9V0S3 PRELIMINARY; PRT; 401 AA.
AC Q9V0S3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN (DFF).
GN PAB1897.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248285; CAB49630.1.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Complete proteome.
SQ SEQUENCE 401 AA; 44290 MW; 96CCD5A190C1FA5D CRC64;

Query Match 14.9%; Score 60; DB 1; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.1e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 IEHVLAGEHENKADLILVCPATANTISKIAGGIDDTPTVVTTVTFPHIPIMIAPAMHE 139
DB 81 IEHVLAGEHENKADLILVCPATANTISKIAGGIDDTPTVVTTVTFPHIPIMIAPAMHE 140

RESULT 3
Q9X213 PRELIMINARY; PRT; 394 AA.
AC Q9X213;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN.
GN TM1687.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS8 / DSM 3109;
RX MEDLINE=95287316; PubMed=10360571;
RA Nelson K.E., Clayton K.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 395:323-329(1999).
DR EMBL; AE001809; AAD36754.1; -.
DR HSSP; Q9SWE5; IE20.
DR TIGR; TM1687; -.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 43804 MW; EBCC811C151ECFA CRC64;

Query Match 3.0%; Score 12; DB 2; Length 394;

Db 218 SSFGKMGVALA 227
DB 208 SSFGKMGVALA 217

Query Match 2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSFGKMGVALA 227
DB 208 SSFGKMGVALA 217

RESULT 4
Q9JYB7 PRELIMINARY; PRT; 394 AA.
AC Q9JYB7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN.
GN NMB1658.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittoni H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002516; AAF42007.1; -.
DR HSSP; Q9SWE5; IE20.
DR TIGR; NMB1658; -.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Complete proteome.
SQ SEQUENCE 394 AA; 42212 MW; B427D1901932AB91 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 SSFGKMGVALA 227
DB 208 SSFGKMGVALA 217

RESULT 5
Q9JYB7 PRELIMINARY; PRT; 394 AA.
AC Q9JYB7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN NMA1916.
GN NMA1916.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=635699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,

```


RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of *Neisseria*
 RT *meningitidis* Z2491.",
 RL Nature 404:502-506(2000).
 DR EMBL; AL162757; CAB85137.1; -.
 DR HSSP; Q9SWES5; 1E20.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 394 AA; 42238 MW; E19917C24E62B087 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 394;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 SSGKMGVALA 227
 Db 208 SSGKMGVALA 217
 |||||

RESULT 6

Q9RQH7 ID Q9RQH7 PRELIMINARY; PRT; 399 AA.
 AC Q9RQH7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG.
 GN DFP.
 OS *Listeria monocytogenes*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC *Bacillus/Staphylococcus* group; *Listeria*.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD;
 RA Milohanic E., Pron B., Boumaila C., Berche P., Gaillard J.L.;
 RT "Isolation and characterization of *Listeria monocytogenes* mutants
 RT altered for adherence to eucaryotic cells.",
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF104226; AAF04763.1; -.
 DR HSSP; Q9SWES5; 1E20.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 SQ SEQUENCE 399 AA; 43395 MW; 566E98FF6F07F7EB CRC64;

Query Match 2.5%; Score 10; DB 2; Length 399;
 Best Local Similarity 100.0%; Pred. No. 0.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 VLVTAGATRE 206
 Db 190 VLVTAGATRE 199
 |||||

RESULT 7

Q9HH70 ID Q9HH70 PRELIMINARY; PRT; 382 AA.
 AC Q9HH70;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE MTW1216.
 GN MTW1216.
 OS *Methanobacterium wolfei*.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC *Methanothermobacter*.
 OX NCBI_TaxID=145261;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Luo Y., Pfister P., Leisinger T., Wasserfallen A.;
 RT "The genome of archaeal prophage psiM100 encodes the lytic enzyme
 RT responsible for autolysis of *Methanothermobacter wolfei*.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF301375; AAG39941.1; -.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 SQ SEQUENCE 382 AA; 41207 MW; 8E792042DFDFA3AB CRC64;

Query Match 2.2%; Score 9; DB 1; Length 382;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 282 AAASVDFRP 290
 Db 257 AAASVDFRP 265
 |||||

RESULT 8

O28628 ID O28628 PRELIMINARY; PRT; 404 AA.
 AC O28628;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PANTOTHENATE METABOLISM FLAVOPROTEIN (DFP).
 GN AFI645.
 OS *Archaeoglobus fulgidus*.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC *Archaeoglobus*.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 DR EMBL; AF000989; AAB89597.1; -.
 DR TIGR; AFI645; -.
 DR InterPro; IPR003382; Flavoprotein.
 DR Pfam; PF02441; Flavoprotein; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 404 AA; 44885 MW; E5A2B899C3E6A66F CRC64;

Query Match 2.2%; Score 9; DB 1; Length 404;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 TANTISKIA 110
 Db 103 TANTISKIA 111
 |||||

RESULT 9

Q9XUT2 ID Q9XUT2 PRELIMINARY; PRT; 127 AA.
 AC Q9XUT2;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

```

DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE K08C9.6 PROTEIN.
GN K08C9.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pezodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lennard N.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX SEQUENCE FROM N.A.
RP MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Alnsough R., Anderson K., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear P., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten T., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.; III of C.
ET "2.2 kb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
DR EMBL; Z81567; CAB04589.1; -.
SQ SEQUENCE 127 AA; 14638 MW; B169A1E75006C2CE CRC64;

Query Match 2.0%; Score 8; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 248 FRIRIKL 255
Db 85 FRIRIKL 92

RESULT 10
Q9RIAL PRELIMINARY; PRT; 142 AA.
AC Q9RIAL;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE SODIUM-HYDROGEN EXCHANGER REGULATORY FACTOR (FRAGMENT).
GN SLC9A3R1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Weinman E.J., Steplock D.A., Zhang X., Akhter S., Shenolikar S.;
RT "Molecular cloning of the cDNA and promoter sequences for the mouse
RT sodium-hydrogen exchanger regulatory factor.";
RL Biochim. Biophys. Acta 0:0-0(1999).
DR EMBL; AF154912; AAD49224.1; -.
DR HSSP; Q12923; 3PDZ.
DR MGD; MGI:1349482; SLC9a3r1.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 142
SQ SEQUENCE 142 AA; 15376 MW; A8994D6A865B283A CRC64;

Query Match 2.0%; Score 8; DB 11; Length 142;
Best Local Similarity 100.0%; Pred. No. 11;

```

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 152 ERLKKLGV 159
Db 97 ERLKKLGV 104

RESULT 11
Q9ZSW7 PRELIMINARY; PRT; 169 AA.
AC Q9ZSW7;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE BETA-FRUCTOFURANOSIDASE (EC 3.2.1.26) (FRAGMENT).
OS Hamamelis virginiana.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Saxifragales; Hamamelidaceae; Hamamelis.
OX NCBI_TaxID=4397;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=LEAF;
RA Szpata M.L.F.;
RT "Isolation of invertase gene sequences from witch hazel (Hamamelis
RL virginiana).";
RL Thesis (1998); Pennsylvania State University.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING BETA-D-
CC -!- FRUCTOFURANOSIDE RESIDUES IN BETA-D-FRUCTOFURANOSIDES.
CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.
DR EMBL; AF091547; AAC99430.1; -.
DR Mendel; 36382; Hamvi;1002;36382.
DR InterPro; IPR001362; Glyco_hydro_32.
DR Pfam; PF00251; Glyco_hydro_32; 1.
DR KEGG; Glycopol; Glycosidase; Hydrolase.
FT NON_TER 1
FT NON_TER 169
SQ SEQUENCE 169 AA; 19268 MW; B74A1632C0D3AF60 CRC64;

Query Match 2.0%; Score 8; DB 10; Length 169;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 LPTGNPVI 73
Db 52 LPTGNPVI 59

RESULT 12
Q9I596 PRELIMINARY; PRT; 210 AA.
AC Q9I596;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HIGH MOBILITY GROUP PROTEIN-1.
GN HMG-1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96174815; PubMed=8599938;
RA Knightingale K., Dimitrov S., Reeves R., Wolfe A.P.;
RT "Evidence for a shared structural role for HMG1 and linker histones B4
RT and H1 in organizing chromatin.";
RL EMBL; J15-548-561(1996).
DR EMBL; U21933; AAC59859.1; -.
DR HSSP; P07155; IAAB
DR InterPro; IPR000135; Highmobility_12.

```

DR InterPro; IPR000910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 2.
 DR PRINTS; PR00886; HIGHMOBLTY12.
 DR SMART; SM00398; HMG; 2.
 DR PROSITE; PS00353; HMG1.2; 1.
 KW Chromosomal protein; DNA-binding; Nuclear protein.
 SQ SEQUENCE 210 AA; 24442 MW; 192CD46D694FF447 CRC64;

Query Match 2.0%; Score 8; DB 13; Length 210;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 286 SDFRPKIK 293
 |||||
 Db 106 SDFRPKIK 113

RESULT 13

Q9CKU9 PRELIMINARY; PRT; 227 AA.
 AC Q9CKU9;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE GPWA.
 GN GPWA OR PM1506.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL; AE006187; AAK03590.1; -.
 DR InterPro; IPR001345; PG_mutase.
 DR Pfam; PF00300; PGAM; 1.
 KW Complete proteome.
 SQ SEQUENCE 227 AA; 25954 MW; B0928CC77B5C6A3 CRC64;

Query Match 2.0%; Score 8; DB 2; Length 227;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 194 GKRVLVTA 201
 |||||
 Db 173 GKRVLVTA 180

RESULT 14

P70441 PRELIMINARY; PRT; 355 AA.
 AC P70441;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PROTEIN CO-FACTOR.
 GN SLC9A3R1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57 BLACK/6; TISSUE=KIDNEY;
 RA Weinman E.J., Stepien D.A., Shenolikar S.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U74079; AAB17569.1; -.

DR HSSP; Q12923; 3PDZ.
 DR MGI; I349482; Slc9a3rl.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 2.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 SQ SEQUENCE 355 AA; 38600 MW; 331FGBEE31DA0A11 CRC64;

Query Match 2.0%; Score 8; DB 11; Length 355;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 152 ERLKKLGV 159
 |||||
 Db 97 ERLKKLGV 104

RESULT 15

Q9X8S5 PRELIMINARY; PRT; 396 AA.
 AC Q9X8S5;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE ASPARTATE AMINOTRANSFERASE.
 GN SCH10.36.
 OS Streptomyces coelicolor.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Murphy L., Harris D.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA James K.P., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL049754; CAB42045.1; -.
 DR HSSP; Q56232; LBKG.
 DR InterPro; IPR001176; ACC_synthase.
 DR InterPro; IPR001511; Aminotran.1.
 DR Pfam; PF00155; aminotran.1.
 DR PRINTS; PR00753; ACCSYNTHASE.
 KW Transferase; Aminotransferase.
 SQ SEQUENCE 396 AA; 42785 MW; 727DBF14C347EBDE CRC64;

Query Match 2.0%; Score 8; DB 2; Length 396;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 VLVTAGAT 204
 |||||
 Db 99 VLVTAGAT 106

Search completed: January 31, 2002, 13:37:33

Job time: 158 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:06 ; Search time 78.64 Seconds
(without alignments)
151.109 Million cell updates/sec

Title: US-08-957-709-71
Perfect score: 156
Sequence: 1 MLLPDWKIRKEILIEPFSE.....PYRGNQGSRLAFSKRKL 156

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	23.7	156	2 E71216	dCTP deaminase (EC
2	22	14.1	134	2 G75030	dCTP deaminase (EC
3	8	5.1	420	2 D84965	dihydrolipoamide S
4	7	4.5	133	1 I40398	flagellar protein
5	7	4.5	133	2 E84101	flagellar protein
6	7	4.5	178	2 A35123	ail protein precur
7	7	4.5	215	2 S74033	alkyl hydroperoxid
8	7	4.5	237	2 C64491	phosphoribosylform
9	7	4.5	268	1 A69000	conserved hypothet
10	7	4.5	279	2 A64149	hypothetical prote
11	7	4.5	327	2 A70435	cysteine synthase
12	7	4.5	362	2 C81445	probable transmemb
13	7	4.5	376	2 T40673	homoserine dehydro
14	7	4.5	439	2 I39524	probable porin - A
15	7	4.5	460	1 S48489	allantoinease (EC 3
16	7	4.5	587	2 H83748	two-component sens
17	7	4.5	661	2 A69252	3-hydroxyacyl-CoA
18	7	4.5	666	2 D82386	methyl-accepting c
19	7	4.5	682	1 G48563	I8 protein - fowlp
20	7	4.5	684	2 T30852	probable RNA helic
21	7	4.5	782	2 D81281	probable nucleotid
22	7	4.5	874	2 H86167	hypothetical prote
23	7	4.5	1017	2 PC4035	cell-cycle-depende
24	7	4.5	1113	2 E64215	hypothetical prote
25	7	4.5	1921	2 T13827	kinesin-73 - fruit
26	6	3.8	34	2 S77846	hypothetical prote
27	6	3.8	35	2 C70256	hypothetical prote
28	6	3.8	37	1 R5EG36	ribosomal protein
29	6	3.8	55	2 H70228	hypothetical prote

30	6	3.8	62	2 I39622	ribosomal protein
31	6	3.8	62	2 H84083	hypothetical prote
32	6	3.8	64	2 T29319	hypothetical prote
33	6	3.8	68	2 G82600	hypothetical prote
34	6	3.8	94	1 VUWTEM	embryonic abundant
35	6	3.8	96	2 B82436	conserved hypothet
36	6	3.8	100	2 G83729	hypothetical prote
37	6	3.8	105	2 S62334	L71-2 protein - fr
38	6	3.8	107	2 T21642	hypothetical prote
39	6	3.8	107	2 T24886	hypothetical prote
40	6	3.8	109	2 A84659	NADS-box protein (
41	6	3.8	110	2 S47555	CI4B9.1 protein -
42	6	3.8	114	1 QXLL3M	NADH dehydrogenase
43	6	3.8	117	2 H72655	protein secretion
44	6	3.8	118	2 T27846	hypothetical prote
45	6	3.8	124	2 A53055	H+-transporting AT

ALIGNMENTS

RESULT 1

E71216
dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: E71216
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hailkawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137
A:Accession: E71216
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-156 <RAW>
A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31124.1; PID:g3258441
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH1997
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 23.7%; Score 37; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 8.5e-31;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAfv 37

Db 1 MLLPDWKIRKEILIEPFSEESLQAGYDLRVGREAfv 37

RESULT 2

G75030
dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75030
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: G75030
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50685.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: dcd; PAB1164
C:Superfamily: dCTP deaminase

C;Keywords: hydrolase

Query Match 14.1%; Score 22; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 3.4e-15;
Matches 22; Conservative 0; Mismatches 0; Indels

Qy	89	GSFAWVDPGWDGNLT	MLYNAS	110
D _b	89	GSFAWVDPGWDGNLT	MLYNAS	110

RESULT 3
D84965
dhidrylipoamide S-succinyltransferase [EC 2.3.1.61] [imported] - Buchnera sp. (strain A
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: D84965
C:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A:Reference number: A84930; MUID:20445173
A:Accession: D84965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: sucB; BU303
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
C:Keywords: acyltransferase; coenzyme A

Query Match	5.13:	Score 8;	DB 2;	Length 420;
Best Local Similarity	100.0%;	Pred. No. 3;		
Matches	8;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
				0;

Qy 45 EKEGKVI 52
 |||||
Db 62 EKEGKVI 69

RESULT 4

I40398

flagellar protein flis - Bacillus subtilis

C:Species: Bacillus subtilis

C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: I40398; E59625

R:Chen, L.; Helmann, J.D.

J. Bacteriol. 176, 3093-3101, 1994

A:title: The Bacillus subtilis sigma D-dependent operon encoding the flagellar proteins

A:Reference number: I40396; MUID:94252974

A:Accession: I40398

A:Status: preliminary; translated from CB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-133 <RES>

A:Cross-references: EMBL:Z31376; RID:g499379; PID:g499382

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, C.; Bron, S.; Broullet, S.; Bruschini, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gallenbach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinios, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mavecel, Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, tuki, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

QY 86 GVIGSFA 92
|||||||
Db 65 GVIGSFA 71

RESULT 7
S74033
A:Accession: S74033
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-215 <SEN>
A:Cross-references: EMBL:Y08256; NID:gl707679; PIDN:CAA69447.1; PID:gl707727
A:Experimental source: strain P2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996
C:Superfamily: alkyl hydroperoxide reductase C22 protein; alkyl hydroperoxidase c22 protein

Query Match 4.5%; Score 7; DB 2; Length 215;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 65 RIKLPDD 71
|||||||
Db 24 RIKLPDD 30

RESULT 8
C64491
A:Accession: C64491
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-237 <BUL>
A:Cross-references: GB:U67594; GB:L77117; NID:gl592160; PIDN:AAB99553.1; PID:gl592163; T
C:Superfamily: N-(5'-phospho-D-riboseylformimino)-5-amino-1-(5''-phosphoribosyl)-4-imidazole

Query Match 4.5%; Score 7; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 46 KEGKVI 52
|||||||
Db 131 KEGKVI 137

RESULT 9
A69000
A:Accession: A69000
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
C:Species: Methanobacterium thermoautotrophicum (strain Delta
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A69000

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
; Olu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: A69000
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <MTH>
A:Cross-references: GB:AE000795; GB:AE000666; NID:g2621036; PIDN:AAB84510.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1
C:Superfamily: conserved hypothetical protein MTH1

Query Match 4.5%; Score 7; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 KEILIEP 16
|||||||
Db 149 KEILIEP 155

RESULT 10
A64149
A:Accession: A64149
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C:Accession: A64149
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: A64149
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-279 <IGR>
A:Cross-references: GB:U32719; GB:L42023; NID:gl573310; PIDN:AAC22006.1; PID:gl573315
A:Note: best homolog was a hypothetical protein from Escherichia coli
C:Superfamily: conserved hypothetical protein HI0345; ferredoxin 2[4Fe-4S] homology
F; 65-130/Domain: ferredoxin 2[4Fe-4S] homology <FER6>

Query Match 4.5%; Score 7; DB 2; Length 279;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 81 SLAREGV 87
|||||||
Db 48 SLAREGV 54

RESULT 11
A70435
A:Accession: A70435
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Jun-2000
C:Accession: A70435
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70435
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-327 <AQF>

A: Cross-references: GB:AE000745; NID:g2983907; PIDN:AAC07459.1; PID:g2983909; GB:AE000606
C: Experimental source: strain VF5
C: Genetics:
A: Gene: *CysM*
C: Superfamily: threonine dehydratase
C: Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate
F: 65/Binding site: pyridoxal phosphate (*Lys*) (covalent) #status predicted

Query Match 4.5%; Score 7; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 KEGKVI 52
|||||
DB 83 KEGKVI 89

RESULT 12
CB1445
probable transmembrane protein Cj0268c [imported] - Campylobacter jejuni (strain NCTC 11168)
C: Species: Campylobacter jejuni
C: Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C: Accession: CB1445
R: Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Whitehead, S.; Barrell, A.G.
Nature 403, 665-668, 2000
A: Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence
A: Reference number: AB1250; UID:20150912
A: Accession: CB1445
A: Status: Preliminary
A: Molecule type: DNA
A: Residues: 1-362 <PAR>
A: Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72736.1; PID:g6967744
C: Experimental source: serotype OZ, strain NCTC 11168
C: Genetics:
A: Gene: Cj0268c

Query Match 4.5%; Score 7; DB 2; Length 362;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 NEPVELR 117
|||||
DB 217 NEPVELR 223

RESULT 13
T40673
homoserine dehydrogenase - fission yeast (Schizosaccharomyces pombe)
C: Species: Schizosaccharomyces pombe
C: Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C: Accession: T40673
R: Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Kutzner, M.; Wambutt, R.
submitted to the EMBL Data Library, January 1999
A: Reference number: Z21944
A: Accession: T40673
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A: Residues: 1-376 <LYN>
A: Cross-references: EMBL:AL035263; PIDN:CAA22876.1; GSPDB:GN00067; SPDB:SPBC776.03
C: Genetics:
A: Gene: SPDB:SPBC776.03
A: Map position: 2
A: Introns: 14/1
C: Superfamily: homoserine dehydrogenase; homoserine dehydrogenase homology

Query Match 4.5%; Score 7; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: January 31, 2002, 13:20:08
Job time: 109 sec



This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:06:08 ; Search time 33.61 Seconds
(without alignments)
888.174 Million cell updates/sec

Title: US-08-957-709-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATKSRKLVGKK.....KMKRELAERIWDEIEKXLS 403

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1985	99.3	403	AAW72844	Polymerase enhanci
2	461.5	23.1	399	AAW72844	S. epidermidis ope
3	444	22.2	422	AAW79946	Corynebacterium gl
4	439	22.0	420	AAW91520	C glutamicum prote
5	197	9.9	228	AAW98418	H. pylori GHPO 319
6	166.5	8.3	204	AAW94179	Human protein sequ
7	165.5	8.3	204	AAW39470	Human polypeptide
8	165.5	8.3	309	AAW41256	Human polypeptide
9	165.5	8.3	326	AAW32199	Human receptor mol
10	162	8.1	217	AAW96816	A. thaliana VB89 (
11	160	8.0	284	AAW93850	Human protein sequ

12	158	7.9	199	21	AAW11382	Arabidopsis thalia
13	158	7.9	209	21	AAW11381	Arabidopsis thalia
14	151.5	7.6	181	20	AAV43439	S. epidermis readi
15	149.5	7.5	181	14	AAW39345	EpD protein. Sta
16	136	6.8	1500	14	AAW30636	hCPSI. Homo sapie
17	135	6.8	31	19	AAW72841	Polymerase enhanci
18	134	6.7	1500	22	AAW49222	Human T1405 carbam
19	134	6.7	1500	22	AAW49223	Human T1405 carbam
20	134	6.7	1500	22	AAW49224	Human carbamyl pho
21	134	6.7	1500	22	AAW49225	Human carbamyl pho
22	129	6.5	35	19	AAW72860	Polymerase enhanci
23	119.5	6.0	568	20	AAW23924	Amino acid sequenc
24	119	6.0	1558	21	AAW18324	Plasmodium falcipa
25	116	5.8	24	19	AAW72863	Polymerase enhanci
26	116	5.8	24	19	AAW72843	Polymerase enhanci
27	111.5	5.6	582	20	AAW23920	Amino acid sequenc
28	109.5	5.5	463	22	AAW90216	C glutamicum prote
29	107	5.4	548	21	AAW01278	Candida glabrata H
30	107	5.4	568	21	AAW01277	Candida glabrata H
31	107	5.4	568	21	AAW01279	Histidine tagged C
32	106	5.3	751	18	AAW13491	Helicobacter ClpB
33	105	5.3	3088	21	AAW19794	Human laminin 2 al
34	105	5.3	3089	21	AAW19792	Human laminin 2 ma
35	105	5.3	3110	16	AAW71730	Merotin major subu
36	105	5.3	3110	20	AAW15460	Human laminin alph
37	105	5.3	3110	21	AAW19791	Human laminin 2 al
38	105	5.3	3110	21	AAW19793	Human laminin 2 al
39	105	5.3	5373	22	AAW14603	Novel bone marrow
40	105	5.3	5447	22	AAW14697	Novel bone marrow
41	103.5	5.2	2541	21	AAW41087	Human ORFX ORF851
42	103	5.2	331	21	AAW04547	Arabidopsis thalia
43	103	5.2	331	21	AAW43327	Arabidopsis thalia
44	102.5	5.1	337	15	AAW63681	Merotin amino acid
45	102.5	5.1	337	18	AAW39206	Human partial mero

ALIGNMENTS

RESULT 1
AAW72844
ID AAW72844 standard; Protein; 403 AA.
XX
AC AAW72844;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor p50 component.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Misc-difference 23 /note= "encoded by NNN"
FT Misc-difference 24 /note= "encoded by NNN"
FT Misc-difference 25 /note= "encoded by NNN"
FT Misc-difference 310 /note= "encoded by NNN"
FT Misc-difference 311 /note= "encoded by NNN"
FT Peptide 1..31 /note= "N-terminal peptide used to generate primers"
FT Peptide 275..291 /note= "internal peptide used to generate primers"
FT Peptide 353..376 /note= "internal peptide used to generate primers"

Db 2 khilavtggiaaykaidtskllqsgdydvrmisdhsaqefvtplafqaisrnpvyntf 61
Qy 75 --EITGFTEHVELAGEHENKADLILVCPATANTISKIACGI-DTPPVTTVTTAPPHIPI 131
Db 62 keepeelqhsvgl-----dwadaliavapatantiaksvgiaddlitslittattpk--- 114
Qy 132 MIAPAMHETMYRHPVIRENRIERKLKLGVEFIGP-----RIEGRKVASIDE 178
Db 115 fvapamnmvnyonprtkhnmkvlsgdgyfyfiegsgylacgyvakgrmee-pmgillsvin 173
Qy 179 IYVRVKKLHKHTLEGRVLYTAGATREYIDPIRIFITNASSGKMGVALAEADPRGA-VT 237
Db 174 kftqkqnvkvsfsgkraltagptveidpvrvysrsgkgayaaealrdkgaivt 233
Qy 238 LIRKGVSKAFIRKIK-LKVTETVEMLSAJENELRSKKYDWMIAAAVSDFRPRIKAE 296
Db 234 li--sgpthslpeginvvkvesadmdfqavte--rfakqdivikaavsytpmdilleh 289
Qy 297 KIK---SGRSITIELVXPXNPIIDRIKEIQPNVLVGPKAETSKELIEGKRQIERAKA 353
Db 290 kkkqegglsvgfk---rtkdilkylgenkthqylvgfaaet--qnieqyaldklkrkna 344
Qy 354 DLV----VGNPLEAFSGENQVLLGRDFTK-ELPKMKKRELAERIWDEIE 399
Db 345 dvilsnvvgdtslgsfssddneitmhfknekvnikkkgkksalahqilelle 395

RESULT 3
AAB79946
ID AAB79946 standard; Protein; 422 AA.
XX AC AAB79946;
XX DT 30-APR-2001 (first entry)
XX DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:626.
KW KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX OS Corynebacterium glutamicum.
XX PN WO200100843-A2.
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000WO-IB00923.
XX PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.

PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX (BADI) BASF AG.
PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
PI WPI; 2001-137957/14.
XX N-PSDB; AAF72065.
PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX Claim 20; Page 1044-1045; 1737pp; English.
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX Sequence 422 AA;
SQ

Query Match 22.28; Score 444; DB 22; Length 422;
Best Local Similarity 31.78; Pred. No. 1.3e-34;
Matches 132; Conservative 85; Mismatches 146; Indels 54; Gaps 20;

Qy 19 KKIVXXXPGSIAALDVKACEGLIR---HGAEEHVAHVSSEATKIIHPYAWNLPTGNPVIT 74
Db 13 invvvgvaggiaa--ykach-ivrafkeagdnrvrvvptesalkfvvgkatfealsgnpvst 69
Qy 75 EI---TGFIEHVELAGEHENKADLILVCPATANTISKIACGIDD--TPVTVVTTAFPHI 129
Db 70 tvfdavdsqvhkvqge----adliviapatadimrvvaglgddllaatlilvatc--- 121

QY 130 PIMTAPAMHETMYRHPVIRENIERKLGVEFGP---RI---EEGRAKVASIDEIVRV 183
 DB 122 pviapamtemwfnpatvanvatlrgrgitviepahgrltgkdtgprlpdpeqiv-dl 180
 QY 184 IKKLH-----KKTLEGRVLTAGATREYIDPIRITNASSGKMGVALAEADFRGA-VT 237
 DB 181 anavhagarlpqdlagkklvltagthheidpvrfignssgrqgfalgeiaaqrghvs 240
 QY 238 LIRTKGSVKAFRIKIKLVETVEEMLSAIENELRSKDYVVIMAAVSDFRPKIKAEKG 297
 DB 241 ivagnaaelptpagaelvppvstqdmfdavqe--ragqsdflvmaaaavadftpasqatsk 298
 QY 298 IKSGRS-----ITIELVPXNPKII-DRIK-----EIOPNVFLVGFKAETSKE--KLIEE 343
 DB 299 lkkgdsdedalstislvenpdilattvkrraeageldsnvpvlgfaaetgdehtale 357
 QY 344 GKROTERAKADLVVGNLT---EAFGSEENQ-VVLIGRDTFKELPKMKKRELAERWD 396
 DB 358 arkkkqkgcdilmcnevngmkvfgqkhnegwildahggvvdvhegskievaagidw 414

 RESULT 4
 ID AAG91520 standard; Protein; 420 AA.
 AC AAG91520;
 DT 26-SEP-2001 (first entry)
 DE C glutamicum protein fragment SEQ ID NO: 5274.
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 OS Corynebacterium glutamicum.
 PN EP1108790-A2.
 PD 20-JUN-2001.
 PF 16-DEC-2000; 2000EP-0127688.
 PR 07-APR-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI: 2001-376931/40.
 XX N-PSDB: AAH66739.
 XX Novel polynucleotides derived from Corynebacterium, for identifying
 XX mutation point of a gene, measuring expression of a gene, analysing
 XX expression profile or pattern of a gene and identifying homologous gene
 XX
 PS Claim 17; SEQ ID NO: 5274; 246pp + Sequence Listing; English.
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from corynebacterium, and identifying a homologue of a gene derived
 CC from corynebacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 SQ Sequence 420 AA;

 Query Match 22.0%; Score 439; DB 22; Length 420;
 Best Local Similarity 31.7%; Pred. No. 4e-34;
 Matches 132; Conservative 84; Mismatches 147; Indels 54; Gaps 20;

 QY 19 KKVXXXPGSTAAALDVKACEGLIR---HGAEVHAMSEAAKIIHPYAWNLPCTGNPVIT 74
 DB 11 rnvvvgvaggiaa--ykach-ivrafkeagdnrvrvptesalkfvgkatfealsgnpvt 67
 QY 75 EI---TGIEHVEVLAAGEHENKADLILVCPANTANTISKIACGIDG--TPVTVVTTAFPHI 129
 DB 68 tvfdavsdvqhvkvgqe---adliaviapataadlmrvvagrddllaatlivac---- 119
 QY 130 PIMTAPAMHETMYRHPVIRENIERKLGVEFGP---RI---EEGRAKVASIDEIVRV 183
 DB 120 pviapamtemwfnpatvanvatlrgrgitviepahgrltgkdtgprlpdpeqiv-dl 178
 QY 184 IKKLH-----KKTLEGRVLTAGATREYIDPIRITNASSGKMGVALAEADFRGA-VT 237
 DB 179 anavhagarlpqdlagkklvltagthheidpvrfignssgrqgfalgeiaaqrghvs 238
 QY 238 LIRTKGSVKAFRIKIKLVETVEEMLSAIENELRSKDYVVIMAAVSDFRPKIKAEKG 297
 DB 239 ivagnaaelptpagaelvppvstqdmfdavqe--ragqsdflvmaaaavadftpasqatsk 296
 QY 298 IKSGRS-----ITIELVPXNPKII-DRIK-----EIOPNVFLVGFKAETSKE--KLIEE 343
 DB 297 lkkgdsdedalstislvenpdilattvkrraeageldsnvpvlgfaaetgdehtale 355
 QY 344 GKROTERAKADLVVGNLT---EAFGSEENQ-VVLIGRDTFKELPKMKKRELAERWD 396
 DB 356 arkkkqkgcdilmcnevngmkvfgqkhnegwildahggvvdvhegskievaagidw 412

 RESULT 5
 ID AAW98418 standard; Protein; 228 AA.
 AC AAW98418;
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPO 319 protein.
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 OS Helicobacter pylori.
 PN WO9843478-A1.
 PD 08-OCT-1998.
 PF 01-APR-1998; 98WO-US06371.
 PR 29-JUL-1997; 97US-0902615.
 PR 01-APR-1997; 97US-0833457.
 PR 24-JUN-1997; 97US-0881227.
 XX (HUMA) HUMAN GENOME SCI INC.
 XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
 XX WPI: 1998-542293/46.
 XX N-PSDB: AAX14137.
 XX New isolated Helicobacter polynucleotides - used to develop products

PT for the diagnosis, prevention and treatment of Helicobacter
XX infections and gastrointestinal diseases
PS Claim 8; Page 700-701; 2054pp; English.

CC This sequence represents a Helicobacter pylori GUPD protein of the
XX invention. The polypeptides can be used for preventing or treating
XX Helicobacter infections, and gastroduodenal diseases associated with
XX these infections, including acute, chronic, and atrophic gastritis, and
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
XX used for the production of antibodies. The products can also be used for
XX detection and diagnosis.

XX Sequence 228 AA;

Query Match 9.9%; Score 197; DB 19; Length 228;
Best Local Similarity 28.3%; Pred. No. 5.6e-11;
Matches 63; Conservative 44; Mismatches 88; Indels 28; Gaps 7;

QY 14 RLVGKRVIVXXPGSIAAL-DVKACEGLIRHGAEVHVMSEAAATKIHPYAWNLPTGNPV 72

DB 12 rllenkrrvlllvsqiaaykslelrvllfksgasqvvmakgkfkplsfalshkv 71

QY 73 ITETGFIHVELAGERH-----KADLILVCPATANTISKIAGIDTPVT-T 120

DB 72 lhd-----rnekwyynhqnalhhnhiacaanadlilfaplstnskskiahadnivs 126

QY 121 VVTTAFPHIPIMTAPAMHETMYRHPVIRENTERLKLGVFEIGPR-----IEEGRAKVA 174

DB 127 fiacaspi---lpsmntmlnsptqsnlkrldksnhiildtknallacdkgdgama 183

QY 175 SIDIVYRVVKKLHKKT-LEGKRVLTAGATREYIDPIREITN 216

DB 184 eptellkkaaqtlkdayfenrevivmgasiekidsvrtisn 226

RESULT 6

AAB94179

ID AAB94179 standard; Protein; 204 AA.

AC AAB94179;

XX 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:14492.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesising polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 14492; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

XX Sequence 204 AA;

Query Match 8.3%; Score 166.5; DB 22; Length 204;
Best Local Similarity 28.1%; Pred. No. 4.4e-08;
Matches 50; Conservative 30; Mismatches 65; Indels 33; Gaps 7;

QY 27 GSTAALDVKACEGLIRH--CAEVHVMSEAAATKIHPY-----AWNL--PTGNP 71

DB 26 gsvaalkpllvsklildipglevsvvttterakfhfyspqdipvtlysdadewmksrds 85

QY 72 VITEITGFIHVELAGERHAKADLILVCPATANTISKIAGIDTPVTVTVTAFPHIPI 131

DB 86 vi-----hidl-----rrwadlllvapldantlkgvasgicdnlitcvmrdrskpl 133

QY 132 MIAPAMHETMYRHPVIRENTERLKLGV---VEFTGPR---EEGRAKVASIDEIVRV 183

DB 134 lfcpanmtamwepitaqvdqkafgyveipcvakklvcgdeglgamaevgtivdkv 191

RESULT 7

AAM39470

ID AAM39470 standard; Protein; 204 AA.

XX AAM39470;

XX 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 2615.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

```

PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0558042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0633450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue A, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX N-PSDB; AAI58626.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2615; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 204 AA;
XX
XX Query Match 8.3%; Score 165.5; DB 22; Length 204;
XX Best Local Similarity 28.1%; Pred. No. 5.5e-08;
XX Matches 50; Conservative 30; Mismatches 65; Indels 33; Gaps 7;
XX
XX QY 27 GSIAALDVKACEGLIRH--GAEVHAVMSEAAATKIHPY-----AWNL--PTGNP 71
XX Db 26 gsvaalkplllvskldipglevavvterakfhfyspqdipvtlysdadewemkrsdp 85
XX
XX QY 72 VITEITGFIEHVELAGHENKADILVCPATANTISKIACGIDTPTVTVTTFAPPHIPI 131
XX Db 86 vl-----hidl-----rrwadlllvapldantlglkvasgicdnlltvcvrawdrskpl 133
XX
XX QY 132 MIAPAMHETWYRHPIVRENTERLKKLG---VEFTGPRI---EEGRKAVASIDEIVYRV 183
XX Db 134 lfcpamtamwhepitaqqvdkafgyveipcvakklvcgdeglgamaevgtivdkv 191
XX
XX RESULT 8
XX ID AAM41256 standard; Protein; 309 AA.
XX
XX AC AAM41256;
XX
XX DT 22-OCT-2001 (first entry)
XX
XX DE Human polypeptide SEQ ID NO 6187.
XX
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

```

```

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX OS Homo sapiens.
XX
XX PN WO200153312-A1.
XX
XX PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000WO-US34263.
XX
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue A, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX N-PSDB; AAI60412.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6187; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 309 AA;
XX
XX Query Match 8.3%; Score 165.5; DB 22; Length 309;
XX Best Local Similarity 28.1%; Pred. No. 1e-07;
XX Matches 50; Conservative 30; Mismatches 65; Indels 33; Gaps 7;
XX
XX QY 27 GSIAALDVKACEGLIRH--GAEVHAVMSEAAATKIHPY-----AWNL--PTGNP 71
XX Db 61 gsvaalkplllvskldipglevavvterakfhfyspqdipvtlysdadewemkrsdp 120
XX
XX QY 72 VITEITGFIEHVELAGHENKADILVCPATANTISKIACGIDTPTVTVTTFAPPHIPI 131
XX Db 121 vl-----hidl-----rrwadlllvapldantlglkvasgicdnlltvcvrawdrskpl 168
XX
XX QY 132 MIAPAMHETWYRHPIVRENTERLKKLG---VEFTGPRI---EEGRKAVASIDEIVYRV 183
XX Db 169 lfcpamtamwhepitaqqvdkafgyveipcvakklvcgdeglgamaevgtivdkv 226
XX
XX RESULT 9
XX AAY32199

```


ID AAY32199 standard; Protein: 326 AA.
 XX AAY32199;
 XX
 DT 15-FEB-2000 (first entry)
 DE Human receptor molecule (REC) encoded by Incyte clone 2022379.
 DE
 KW Receptor; REC; human; diagnosis; therapy; neoplastic disorder;
 KW immunological disorder; reproductive disorder; nervous disorder;
 KW gastrointestinal disorder; smooth muscle disorder;
 KW musculoskeletal disorder.
 XX
 OS Homo sapiens.
 XX
 PN W09957270-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 28-APR-1999; 99WO-US09191.
 XX
 PR 01-MAY-1998; 98US-0071822.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Hillman JL, Bandman O, Tang YT, Yue H, Lal P, Corley NC;
 PI Guegler KJ, Patterson C;
 XX
 XX WPI: 2000-052971/04.
 DR N-PSDB; AAZ34608.
 XX
 PT Novel human receptor molecules used in the diagnosis, treatment and
 PT prevention of neoplastic, immunological, reproductive gastrointestinal,
 PT nervous, smooth muscle and musculoskeletal disorders
 XX
 PS Claim 1; Page 71-72; 94pp; English.
 XX
 CC The present sequence represents a human receptor molecule (REC)
 CC encoded by Incyte cDNA clone 2022379. The invention provides human
 CC RECs and polynucleotides which identify and encode REC, as well as
 CC vectors, host cells, antibodies, agonists and antagonists. Human
 CC RECs appear to play a role in neoplastic, immunological,
 CC reproductive gastrointestinal, nervous, smooth muscle and
 CC musculoskeletal disorders. The protein, antagonists and agonists,
 CC and compositions can be used to treat: a reproductive disorder,
 CC including but not limited to, prolactin production disorders,
 CC infertility including tubal disease, ovulatory defects, endometriosis,
 CC disruptions of the oestrous and menstrual cycles, polycystic ovary
 CC syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian
 CC tumours, uterine fibroids, autoimmune disorders, ectopic pregnancy,
 CC teratogenesis, breast cancer, fibrocystic breast disease,
 CC galactorrhoea, disruptions of spermatogenesis, abnormal sperm
 CC physiology, testis cancer, prostate cancer, benign prostatic
 CC hyperplasia, prostatitis, Peyronie's disease, male breast carcinoma
 CC and gynecomastia; gastrointestinal disorders including, but are not
 CC limited to, dysphagia, peptic oesophagitis, oesophageal spasm and
 CC stricture, oesophageal carcinoma, dyspepsia, indigestion, gastritis,
 CC gastric carcinoma, anorexia, nausea, emesis, gastroparesis, intestinal
 CC tract infection, peptic ulcer, colitis, Whipple's disease, Mallory-Weiss
 CC syndrome, irritable bowel syndrome, short bowel syndrome, diarrhoea,
 CC constipation, cirrhosis, jaundice, and hepatic vein thrombosis; nervous
 CC disorders including, but are not limited to, Alzheimer's disease,
 CC amnesia, bipolar disorder, catatonia, cerebral neoplasms, Down's
 CC syndrome, and dystonias; smooth muscle cell disorders including, but
 CC not limited to, angina, anaphylactic shock, arrhythmia, cardiovascular
 CC shock, migraine, and pheochromocytoma; musculoskeletal disorders
 CC including muscular dystrophy, central core disease, nemaline myopathy,
 CC centronuclear myopathy, lipid myopathy, inclusion body myositis,
 CC thyrotoxic myopathy, and ethanol myopathy; immunological disorders
 CC including AIDS, Addison's disease, adult respiratory distress
 CC syndrome, allergy, ankylosing spondylitis, amyloidosis, anaemia,
 CC asthma, atherosclerosis, autoimmune hemolytic anaemia, autoimmune
 CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's

CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,
 CC emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis
 CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,
 CC Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis,
 CC hyperesinophilia, irritable bowel syndrome, multiple sclerosis,
 CC myasthenia gravis, myocardial or pericardial inflammation,
 CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, systemic
 CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's
 CC syndrome, systemic anaphylaxis, systemic lupus erythematosus, psoriasis,
 CC sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis,
 CC Werner syndrome, complications of cancer, haemodialysis, and
 CC extracorporeal circulation; viral, fungal, bacterial or protozoan
 CC infections; trauma; and neoplastic disorders including adenocarcinoma,
 CC leukemia, melanoma, myeloma, sarcoma, and various cancers. The REC
 CC polynucleotide is a source of probes and primers which bind may be used
 CC to detect REC in a sample from a patient (claimed). They may also be
 CC administered as part of a gene therapy regime.
 XX
 SQ Sequence 326 AA;

Query Match 8.3%; Score 165.5; DB 21; Length 326;
 Best Local Similarity 28.1%; Pred. No. 1.1e-07;
 Matches 50; Conservative; 30; Mismatches 65; Indels 33; Gaps 7;
 QY 27 GSIALDVKACEGLIRH--GAEVHVMSEAATKIHPY-----ANWL--PTGNP 71
 DB 26 gsvaalkpllvskldipglevavvttterakhsfydpdipvtlysdadewmksrddp 85
 QY 72 VITEITGFIEHVELAGEHENKADLILVCPATANTISKIACGIDDTPTVTVTTAPPHIPI 131
 DB 86 vl-----hidl-----rrwadlllvapldantlgkvasgicdnlltcmvrawdrskpl 133
 QY 132 MIAPAMHETWTRHPIVRENIERLKLKLG---VEFIGPRT---EEGRKVASIDEIVYRV 193
 DB 134 lfcpmantamwhepitaqqvdklkafgyveipcvakkllvcgdeglgamaevgtivdkv 191

RESULT 10
 AAY96816
 ID AAY96816 standard; Protein: 217 AA.
 XX
 AC AAY96816;
 XX
 DT 26-SEP-2000 (first entry)
 XX
 DE A. thaliana Vb89 (HAL3) CDC2b interacting protein.
 XX
 KW Cyclin-dependent protein kinase; CDK; CDC2a; CDC2b; Vb89; HAL3;
 KW cell cycle; interacting protein; environmental stress; growth regulator;
 KW herbicide; nematode resistance; plant breeding.
 XX
 OS Arabidopsis thaliana.
 XX
 PN W0200036124-A2.
 XX
 PD 22-JUN-2000.
 XX
 PF 17-DEC-1999; 99WO-EPI0084.
 XX
 PR 17-DEC-1998; 98EP-0124062.
 XX
 XX (CROP-) CROPEDESIGN NV.
 PA
 XX De Veylder L, Boudolf VKCK, Torres Acosta JA, Inze D;
 PI
 XX WPI: 2000-431601/37.
 DR N-PSDB; AAA51412.
 XX
 PT Nucleic acids encoding plant cell cycle interacting proteins, useful
 PT for regulating plant growth and in recombinant DNA protocols
 XX
 PS Claim 1; Page 126; 152pp; English.

XX The Vb89 clone encodes the Arabidopsis thaliana HML3 homologue, a
CC halotolerant gene isolated in Saccharomyces cerevisiae. The Vb89 clone
CC interacts with A. thaliana CDC2b (a cyclin-dependent protein kinase
CC (CDK)), but not with CDC2a in the two-hybrid system.
CC CDC2a and CDC2b are the only CDK genes to have been characterized in
CC detail in Arabidopsis thaliana. They were used as bait in a two-hybrid
CC screening assay with a cDNA library of a plant cell suspension as prey.
CC The plant cell cycle interacting proteins identified were designated
CC LbV15, pHO80-like protein, Vb33, Vb89, VbAHP and VbHSF. The nucleic
CC acids, vectors comprising them, the proteins they express, antibodies
CC that bind to them and/or inhibitors of their protein expression and/or
CC activity may be used for modulating the cell cycle in an animal or plant,
CC plant cell division and/or growth, for influencing the activity of cell
CC cycle proteins in a plant or animal cell, as positive or negative
CC regulators of cell proliferation, for modifying the growth inhibition
CC caused by environmental stress conditions (e.g. to improve growth of
CC plants in normal or suboptimal nutrient conditions, especially
CC of cell cycle protein, as growth regulators, herbicides and/or for
CC inducing nematode resistance in plants. The DNA sequences and their
CC regulatory sequences may be used as markers in plant or animal cell
CC tissue cultures or as a marker in marker-assisted plant breeding. The
CC regulatory sequences may also be used for the expression of heterologous
CC DNA sequences during a stage of the cell cycle.
XX Sequence 217 AA;
SQ

Query Match 8.1%; Score 162; DB 21; Length 217;
Best Local Similarity 27.1%; Pred. No. 1.3e-07;
Matches 46; Conservative 33; Mismatches 59; Indels 32; Gaps 5;
Qy 11 TSKRLVKKVXXPGSIAALDKVACGLIRHGAEVHVMSEATKIHP----- 61
Db 23 tvtrk---prillaasgsvaiksfnlchcfsewaevkavaskslnfvdkslpqnvtl 79
Qy 62 -----YANLPTGNPVITEIGTIEHVELAGEHENKADLILVCPATANTISKIACGID 114
Db 80 ytdedewssw-nkgdpyl-----hiel-----rrwadvmilaplsantlaklaglc 126
Qy 115 DTPVTVTYTFAPPHIPMIAPAMHETMYRHPVIRENIERLKKLGVETGTP 164
Db 127 dnlltcivrawdyskplvapamntlmwnpfterhlvldelgitlpp 176

RESULT 11
AAB93850
ID AAB93850 standard; Protein; 284 AA.
XX
AC AAB93850;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13699.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA {HELI-} HELIX RES INST.

XX Ota T. Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 13699; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides; and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX Sequence 284 AA;
SQ

Query Match 8.0%; Score 160; DB 22; Length 284;
Best Local Similarity 26.8%; Pred. No. 3.1e-07;
Matches 78; Conservative 43; Mismatches 92; Indels 78; Gaps 15;
Qy 160 EFIGPRIEGRAKVASIDEIVYRVKIKLHKLTLECKR-VLVTAGATREYID-PPIREITN 216
Db 9 efqap---pgaarwa---evmarfaarl---gagrrvltsgtkvplearpvrldn 59
Qy 217 ASSGKMVGVALAE--ADPRGAVTLIRTKGS-----VKAFR---IRKIKLVE 258
Db 60 fssgrgatsaeafaaaggygvlflrarsafpahrppqtwlsalpsgalgllsie 119
Qy 259 TVEMLSAIENELRSKK-----YDVVIMAA 283
Db 120 aeenalpgfaealrsgaeaaagtlavefttladyhlhlaaaqalnplgpsamfyiaa 179
Qy 284 AVSDFRPKIK--AEGKIKSG---RSITIELVPXNPKIID-RIKEIQPNVFLVGFKAETSK 337
Db 180 avsfyfvpsvemphehkiqsggplqitlmkv---pkllsylvkdwakpafklsfkltdp 236
Qy 338 EKLIEEGKRIERAKADLVVGNLTLEAFGSEENQVVLIGRDF-TRELPMKK 387
Db 237 aivnarkaleiyghqvvanile---srqsfvltkdssetklllsrkk 284

RESULT 12
AAG11382
ID AAG11382 standard; Protein; 199 AA.
XX
AC AAG11382;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10077.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

PR 23-JUN-1999;

PR 23-JUN-1999;

PR 24-JUN-1999;

PR 28-JUN-1999;

PR 29-JUN-1999;

PR 30-JUN-1999;

PR 01-JUL-1999;

PR 01-JUL-1999;

PR 02-JUL-1999;

PR 06-JUL-1999;

PR 08-JUL-1999;

PR 09-JUL-1999;

PR 12-JUL-1999;

PR 13-JUL-1999;

PR 14-JUL-1999;

PR 15-JUL-1999;

PR 16-JUL-1999;

PR 16-JUL-1999;

PR 19-JUL-1999;

PR 19-JUL-1999;

PR 19-JUL-1999;

PR 19-JUL-1999;

PR 20-JUL-1999;

PR 20-JUL-1999;

PR 20-JUL-1999;

PR 21-JUL-1999;

PR 21-JUL-1999;

PR 22-JUL-1999;

PR 22-JUL-1999;

PR 22-JUL-1999;

PR 22-JUL-1999;

PR 23-JUL-1999;

PR 23-JUL-1999;

PR 23-JUL-1999;

PR 26-JUL-1999;

PR 26-JUL-1999;

PR 27-JUL-1999;

PR 27-JUL-1999;

PR 28-JUL-1999;

PR 02-AUG-1999;

PR 02-AUG-1999;

PR 02-AUG-1999;

PR 03-AUG-1999;

PR 04-AUG-1999;

PR 04-AUG-1999;

PR 05-AUG-1999;

PR 05-AUG-1999;

PR 06-AUG-1999;

PR 06-AUG-1999;

PR 09-AUG-1999;

PR 10-AUG-1999;

PR 11-AUG-1999;

PR 12-AUG-1999;

PR 13-AUG-1999;

PR 13-AUG-1999;

PR 16-AUG-1999;

PR 17-AUG-1999;

PR 18-AUG-1999;

PR 20-AUG-1999;

PR 23-AUG-1999;

99US-0140353.

99US-0140354.

99US-0140895.

99US-0140823.

99US-0140991.

99US-0141287.

99US-0141842.

99US-0142154.

99US-0142055.

99US-0142390.

99US-0142803.

99US-0142920.

99US-0142977.

99US-0143542.

99US-0143624.

99US-0144005.

99US-0144085.

99US-0144086.

99US-0144325.

99US-0144331.

99US-0144332.

99US-0144333.

99US-0144334.

99US-0144335.

99US-0144352.

99US-0144632.

99US-0144884.

99US-0144814.

99US-0145086.

99US-0145088.

99US-0145085.

99US-0145087.

99US-0145089.

99US-0145192.

99US-0145145.

99US-0145218.

99US-0145224.

99US-0145276.

99US-0145913.

99US-0145918.

99US-0145919.

99US-0145951.

99US-0146386.

99US-0146388.

99US-0146389.

99US-0147038.

99US-0147204.

99US-0147302.

99US-0147302.

99US-0147260.

99US-0147303.

99US-0147416.

99US-0147493.

99US-0147935.

99US-0148171.

99US-0148319.

99US-0148341.

99US-0148565.

99US-0148684.

99US-0149368.

99US-0149175.

99US-0149426.

99US-0149722.

99US-0149723.

99US-0149929.

99US-0149902.

99US-0149930.

99US-0150566.

99US-0150884.

99US-0151065.

99US-0151066.

99US-0151080.

99US-0151303.

```

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 7.9%; Score 158; DB 21; Length 199;
Best Local Similarity 25.0%; Pred. No. 2.9e-07;
Matches 52; Conservative 44; Mismatches 68; Indels 44; Gaps 10;

Qy 11 TKSRLVGKIVXXPGSTIALDVKACEGLIRHGAENVHVMSEA-----TKIIHPY---- 62
Db 5 ttpk---prillaasgvaakifgnlchcftewaevravvkskshfklkispqevtl 61
Qy 63 -----AWNLPGNPVITETGFIHVELAGHENKADILVCPATANTISKIACGID 114
Db 62 ytdedewssw-kigdpvi-----hiei-----rrwadvivlapiasantigkiaggic. 108
Qy 115 DTPVTVTVTAPPHIPIMAPAMHETWTRHPIVRENIERLKKLGVETFGP---RI---EE 168
Db 109 dnllccilrawdytkplivapamnlcmwnpfterhlisldelgitlippiakrlacgdy 168
Qy 169 GRAKVASIDEIVYRVK-----KLHKKT 191
Db 169 gngamae-psliystvrfwesqhagt 195

RESULT 13
AAG11381
ID AAG11381 standard; protein: 209 AA.
XX

```

```

AC AAG11381;
XX 17-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 10076.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10076.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10076.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EP1033405-A2.
PN 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

```

PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139829.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143342.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	03-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
			Query Match 7.9%; Score 158; DB 21; Length 209;		
			Best Local Similarity 25.0%; Pred. No. 3.le-07;		
			Matches 52; Conservative 44; Mismatches 68; Indels 44; Gaps 10;		
QY	11	TKSRKLVGKKIVXXPGSIAALDVKACEGLIRHGAEVHVMSEA-----TKIHPY----- 62			
Db	15	ttprk---prvllaagsvaaikfnlchcftewaevravvtksslhfdklsldpdevtl 71			
QY	63	-----AWNLTGNPVITEITGTEHVELAGEHENKADLIIVCPATANTISKIACGID 114			
Db	72	ytdedewsswn-kigdpvl-----hiei-----rrwadviaplantsantikiaggic 118			
QY	115	DTPVTTVTTAFPHIPIMIPAMHETMYRPIVRENIERLKKIGVFEFIGP---RI---EE 168			
Db	119	dnlltciirawdytkplfvapamntlmwnpfterhlslldelgitlippiakkracgdy 178			
QY	169	GRAKVASIDEIVRYVIK-----KLHKKT 191			
Db	179	ngamae-psliystvrlfwesqahqgt 205			

RESULT 14
 ID AAY43439 standard; Protein; 181 AA.
 XX AAY43439
 AC AAY43439;
 DT 17-DEC-1999 (first entry)
 XX
 DE S. epidermis reading frame epid protein sequence.
 XX
 KW Staphylococcus epidermis; epidermin; epia; epib; epic; epid; epi;
 KW epig; epi'; epi'; oxidative carboxylation; flavoprotein.
 XX
 OS Staphylococcus epidermis.
 XX
 PN US5962253-A.
 XX
 PD 05-OCT-1999.
 XX
 PF 13-MAY-1996; 96US-0645193.
 XX
 PR 13-MAY-1996; 96US-0645193.
 XX
 PA (BOEH) BOEHRINGER INGELHEIM INT GMBH.
 XX
 PI Kupke T, Kempter C, Jung G, Goetz F;
 XX
 DR WPT; 1999-589714/50.
 XX
 PT Oxidative decarboxylation of peptides by the flavoprotein Epid -
 XX
 PS Claim 1; Fig 9; 89pp; English.
 CC
 CC The present invention describes the oxidative decarboxylation of a
 CC peptide comprising combining the peptide which contains at its carboxy
 CC terminus the amino acid sequence (I) with Epid; X1-X2-X3-X4-X5-X6-C (I)
 CC where X1, X2, X3 and X4 - any one of the 20 common amino acids; X5 -
 CC Tyr, Val, Met, Phe, Ile, Leu or Trp; and X6 - Cys, Ala, Ser, Val or Thr
 CC (2)-(5): SENSICC (2); SENSICC (3); SENSICC (4) and SENSICC (5). The
 CC method can be used for the preparation of derivatives of known peptides
 CC and hormones with a cysteine residue in the unmodified peptide replaced
 CC by sulphur-bridged amino acids and serine and thiamine replaced
 CC corresponding dehydroamino acid residues. The formation of novel
 CC compounds for experimental purposes or for the formation of known
 CC peptide derivatives in new hosts e.g. biologically active
 CC least one lanthionine bridge and/or methyl-lanthionine bridge and/or at
 CC derivatives of human insulin, oxytocin, vasopressin, peptide
 CC antibiotics, hormone inhibitors such as elastase inhibitor and
 CC fibrinolytically active agents such as human tissue plasminogen
 CC activator can also be achieved using this method. Derivatives modified
 CC by this method retain the biological activity of the parent compound but
 CC have increased stability and improved half-lives. AAY43418 to AAY43475,
 CC and AAY25496 to AAY25506, represent sequences used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 181 AA;

Query Match 7.6%; Score 151.5; DB 20; Length 181;
 Best Local Similarity 29.5%; Pred. No. 1.1e-06;
 Matches 56; Conservative 37; Mismatches 76; Indels 21; Gaps 9;
 QY 16 LVGKKIVXXPGSTAAADVAK-ACEGLIRHGAEVHVMSEANAKLIHPYAWNLPNGPVIT 74
 Db 1 mygkllcat-aslnvinhlyvelkqhtdevnlfspsknfntdvklfcfn-lyd 58
 QY 75 EITG-FIEHVELAGEHKNADLIVCPATANTISKACGIDTPVTVTAFPHIPMI 133
 Db 59 eikdplnlhinivenhe---yilvipasantinkiangicdnllttvtcltgyq--kifi 112

QY 134 APAMHETMYRHPIVRENIERLKKLGVEFTGP-----RIEGRK-----VASIDEIVYRV 183
 Db 113 fpmnirmwgnpflqknidllknndkvyspdmnksfessgryknitmpnienvinfv 172
 QY 184 IKKLHKKTLE 193
 Db 173 inn-ekrpid 181
 RESULT 15
 AAR39345
 ID AAR39345 standard; Protein; 181 AA.
 XX AAR39345;
 DT 15-SEP-1993 (first entry)
 XX
 DE Epid protein.
 XX
 KW Epidermin; derivatives; lantibiotic.
 XX
 OS Staphylococcus epidermis.
 XX
 PN EP543195-A.
 XX
 PD 26-MAY-1993.
 XX
 PF 30-OCT-1992; 92EP-0118598.
 XX
 PR 31-OCT-1991; 91US-0784234.
 XX
 PS (THOM) THOMAE GMBH KARL.
 XX
 PI Augustin J, Engelke G, Entian K, Gotz F, Jung G, Kaletta C;
 PI Klein C, Kallner R, Kupke T, Rosenstein R, Schnell N, Wieland B;
 DR WPI; 1993-168917/21.
 DR N-PSDB; AAQ42541.
 XX
 PT Novel DNA molecule - encoding Epi B, C, D, P or Q enzymes
 PT involved in biosynthesis of lantibiotic epidermin.
 XX
 PS Claim 5; Fig 9; 52pp; English.
 CC
 CC The sequence is that of Epid which is involved with EpiB and EpiC
 CC in the four enzymatic modification reactions involved in the prodn.
 CC of epidermin, (1) water elimination by a serine/threonine dehydratase,
 CC (2) sulphur addition by a lanthionine synthase, (3) C-terminal
 CC carboxylation by a cysteine decarboxylase and (4) double bond
 CC formation.
 XX
 SQ Sequence 181 AA;

Query Match 7.5%; Score 149.5; DB 14; Length 181;
 Best Local Similarity 30.9%; Pred. No. 1.7e-06;
 Matches 51; Conservative 31; Mismatches 64; Indels 19; Gaps 7;
 QY 40 LIRHGAEVHVMSEANAKLIHPYAWNLPNGPVITEITG-FIEHVELAGEHKNADLILV 98
 Db 25 lkqhtdevnlfspsknfntdvklfcfn-lydelkdplnlhinivenhe---yilv 79
 QY 99 CPATANTISKACGIDTPVTVTAFPHIPMIAPAMHETMYRHPIVRENIERLKKLG 158
 Db 80 lpsantinkiangicdnllttvtcltgyq--kifipmnmirmwgnpflqknidllknnd 137
 QY 159 VEFTGP-----RIEGRK-----VASIDEIVYRVIKKLHKKTLE 193
 Db 138 kvyspdmnksfessgryknitmpnienvinfvinn-ekrpid 181

Search completed: January 31, 2002, 13:07:05

Job time: 57 sec

This Page Blank (uspto)

This Page Blank (uspto)

Result No.	Query			ID	Description	
	Score	Match	Length			
1	1988	99.5	403	4	US-08-822-774-19	Sequence 19, Appl
2	635.5	31.8	437	4	US-08-822-774-51	Sequence 51, Appl
3	438.5	21.9	444	4	US-08-822-774-52	Sequence 52, Appl
4	151.5	7.6	181	2	US-08-645-193B-24	Sequence 24, Appl
5	149.5	7.5	181	2	US-08-392-625-22	Sequence 22, Appl
6	149.5	7.5	181	2	US-08-466-961A-22	Sequence 22, Appl
7	135	6.8	31	4	US-08-822-774-46	Sequence 46, Appl
8	116	5.8	24	4	US-08-822-774-6	Sequence 6, Appl
9	116	5.8	24	4	US-08-822-774-48	Sequence 48, Appl
10	105	5.3	3111	2	US-08-460-309-4	Sequence 4, Appl
11	105	5.3	3111	2	US-08-126-077-4	Sequence 4, Appl
12	104	5.2	35	4	US-08-822-774-3	Sequence 3, Appl
13	102.5	5.1	337	1	US-08-311-223-3	Sequence 3, Appl
14	102.5	5.1	337	3	US-09-059-849A-3	Sequence 3, Appl
15	102.5	5.1	337	3	US-09-213-632-3	Sequence 3, Appl
16	102.5	5.1	337	5	PCT-US95-12675-3	Sequence 3, Appl
17	101.5	5.1	723	1	US-07-814-964-11	Sequence 11, Appl
18	101.5	5.1	723	1	US-08-258-442-11	Sequence 11, Appl
19	101.5	5.1	723	1	US-08-328-809-6	Sequence 6, Appl
20	101.5	5.1	723	5	PCT-US92-11107-11	Sequence 11, Appl
21	101	5.1	1786	4	US-08-973-462-8	Sequence 8, Appl
22	99.5	5.0	337	1	US-08-445-135-4	Sequence 4, Appl
23	98.5	4.9	1030	4	US-09-091-117-2	Sequence 2, Appl
24	98	4.9	630	4	US-08-973-462-9	Sequence 9, Appl
25	96	4.8	547	2	US-08-467-822-35	Sequence 35, Appl
26	96	4.8	547	4	US-08-432-697-35	Sequence 35, Appl
27	96	4.8	547	4	US-08-466-248-35	Sequence 35, Appl

```
Qy 61 PYAWNLPTGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACIDTPVTT 120
Db 61 PYAWNLPTGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACIDTPVTT 120
Qy 121 VYTTAFPHIPIMAPAMHETMYRHPDIVRENTERLKKLVGVEFGPRIEGRKAVASIDEIV 180
Db 121 VYTTAFPHIPIMAPAMHETMYRHPDIVRENTERLKKLVGVEFGPRIEGRKAVASIDEIV 180
Qy 181 YRVIKLHKHTLEGKRVLVLTAGATREYIDPIRFTINASSSGKMGVALAEADFRGAVTLIR 240
Db 181 YRVIKLHKHTLEGKRVLVLTAGATREYIDPIRFTINASSSGKMGVALAEADFRGAVTLIR 240
Qy 241 TKGSKAFRIKIRKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDPRPKIKABGKIKS 300
Db 241 TKGSKAFRIKIRKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDPRPKIKABGKIKS 300
Qy 301 GRSITIELVPXNPKIIDRIKETOPNVFLVGFKAETSKELTEEGKROIERAKADLVGNT 360
Db 301 GRSITIELVPXNPKIIDRIKETOPNVFLVGFKAETSKELTEEGKROIERAKADLVGNT 360
Qy 361 LEAFGSEENQVVLIGRDTKELPKMKKRELAERIWDEIEKXLS 403
Db 361 LEAFGSEENQVVLIGRDTKELPKMKKRELAERIWDEIEKXLS 403

RESULT 2
US-08-822-774-51
; Sequence 51, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Methanococcus Jannaschii
US-08-822-774-51

+ Query Match 31.8%; Score 635.5; DB 4; Length 437;
```

```
Best Local Similarity 37.9%; Pred. No. 2.3e-52;
Matches 162; Conservative 83; Mismatches 153; Indels 29; Gaps 12;
Qy 1 MLHHVKLIATKSRKLVGKKIVXXXPGSIAALDV-KACEGLIRHGAEVHVMSEAATKII 59
Db 5 IMHPTKLLKGTSKLLENKKILVAVTSSIAAETPKMLRELIRHGAEVYCIITEETKII 64
Qy 60 HPYAWNLPTGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIACID 114
Db 65 GKEALKFCGNEVBEITGXXXXXDIHILLYXXXNXCDCLLIYPATANTISKINL 124
Qy 115 DTPVTVVYTTAFPHIPIMAPAMHETMYRHPDIVRENTERL-KLGVEFGPRIE- 168
Db 125 DNVNTTALMEFGNKPIFIVPAMHENMENXXAIRHIDKLEKDKIYIISPKFEXXXX 184
Qy 169 -GRAKVASIDEIVRVIKKL-HKVTLECKRVLVLTAGATREYIDPIRFTINASSGKMGVAL 226
Db 185 XGKAKVANIEDVRAVIEKIGNNKKEGNRVLILNGGTVEFDKRVVISLSSGKMGVAL 244
Qy 227 AEEADFRG-AVTLIRTKGSVKAFRIKIKLVETVEMLSAIENELRSKKYDVVIMAA 284
Db 245 AEAFCKEGYVEVITAMGLEPPYIK--NHRVLTAKEMLNKAIEXXLAKDFDIISAA 302
Qy 285 VSDFRPKIKABGKIKS-----GRSITIELVPXNPKIIDRIKETIQPNV-FLVGFKAP--TSK 337
Db 303 ISDETVESXFEGLSSSEEXXXLILKLRXNPKVLEELRRIYKDXKVLIGFKAEYNDE 362
Qy 338 EKLIIEGKROIERAKADLVGNTLEA----FGSEENQVVLIGRDTKELPKMKKRELAER 393
Db 363 KELINRAKERLNKYNLNMIIANDLSKXXHYEGDDYIEVYIITKVEIKSGKXKXETISER 422
Qy 394 IWDEIEK 400
Db 423 IVERVKK 429

RESULT 3
US-08-822-774-52
; Sequence 52, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 444 amino acids
; TYPE: amino acid
```

```

; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
US-08-822-774-52

```

Query Match 21.9%; Score 438.5; DB 4; Length 444;
Best Local Similarity 32.9%; Pred. NO. 1.3e-33;
Matches 136; Conservative 70; Mismatches 172; Indels 35; Gaps 14;

QY	16	LVGRKVVXXPGSIAALDYKACBGLTR-HGAEVHVAWSEAAKTIHPYAWNLP	TGNPVT	74
Db	27	LAGKIKVLVGGSGIAAYKTPELVRRLRDGDYRVAAANTEAAKAFITPLSLQAVSGY	PPSD	86
QY	75	EI-----TGFIEHVELAGEHENKADLILVCPATANTISKIACIGDTPVT	VVVVTFAPHI	129
Db	87	SLIDPAEAAAMGHIELGXKXKXWADUILLAPATAJUIARVAAGMNDLVSTICLATP	PPXA	146
QY	130	PIIMAPAMHETRYRHPIVRENIERL-KKLGVEFIPRIEGRKVASI-----DEIVYR	182	
Db	147	PVAVLPA MNQMYRAAATQHNLEVLAXSRGLLWGP-DSGSOACGDIGPRXDXPLTIV	204	
QY	183	VIKKLH---KKTLEGRKVLVTAGATREYIDPRTFTNASSGKMGMVALABEADFRGA-VTL	238	
Db	205	DMAVAHFSPVNDLKLHINIMITAGTPREPLDPVRYISNHSSGKMGFIAAAAAARRRGNVTL	264	
QY	239	IRTKGSKAFRIRKIKLVKETEVEEMLSAENELRSKKYDVTVMMAAVSDFRPIKAEGKI	298	
Db	265	VSGVSLTPPPFVK-RDVMTALEMAAVNXXASVOQQNIFIGCAAVADYRAATVAPEKI	323	
QY	299	K-----SGRSITIELVPXPKI---IDRIKEIOPNVLGVFKAEATSK-EKLEIEGKRQIE-	349	
Db	324	KKQATQGDDELTKWVKXNPDIIVAGVAALKDHRP-VVVGFAAETNNXXXXVEYARQKI	381	
QY	350	RAKADLWGNWL-----EAFGSEENQVVLIGRDTFKELPKMKKRELAERIWDEI	398	
Db	382	RKNLDITCANDVSQPTQGFNSDNNAHLHFWQDGKVLPLERKELLCOLLDEI	434	

```

RESULT      4
US-86-645-193B-24
; Sequence 24, Application US/08645193B
; Patent No. 5962253
; GENERAL INFORMATION:
; APPLICANT: Kupke, Thomas
; APPLICANT: Gotz, Friedrich
; APPLICANT: Kempter, Christoph
; APPLICANT: Jung, Gunther
; TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
; TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/645.193B
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893

```

```

: REFERENCE/DOCKET NUMBER: 0652.1540000
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 24:

```

	Query Match	7.6%; Score 151.5; DB 2;	Length 181;
	Best Local Similarity	29.5%; Pred. No. 7.3e-07;	
	Matches 56; Conservative 37;	Mismatches 76; Indels 21;	Gaps
Qy	16 LVGRKVIXXPGSIAADVK--ACEGLIRHGAEVHAVNSEAATKIHHPYÄNNLTGPNPVIT 74	: : : : : :	:
Db	1 MYGRKKLCAT--ASINVININHYYVELAQHQFDEVNILEPSSKNKFINTDVLKLCDN-LYD 58	: : : : : :	:
Qy	75 EITG-FTEHVELAGEHENKADLLVCAPTANTTSKIIACGIDDDTPVTVVTTAPHPIDIMI 133	: : : : : :	:
Db	59 EIKDPPLLNHNIVENHR-----YILVLPASANTINKIANGICDNLTLTTCVTGYQ--KLFI 112	: : : : : :	:
Qy	134 APAMHEMTYRHPIVRENIERLKKLGVEFGP-----RIEGRAK----VASIDEIVRYR 183	: : : : : :	:
Db	113 FPNWNIRWGNPFLOKNIDLKNDNVKVYSPDNMKSFEISSGRYKKNITMPNIENVLVNFV 172	: : : : : :	:
Qy	184 IKKLHKKTLE 193	:	:
Db	173 LNN-EKRPLD 181	: : : :	:

RESULT 5
US-08-392-625-22
; Sequence 22, Application US/08392625
; Patent No. 5837485
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
; APPLICANT: G tz, Friedrich
; APPLICANT: Schnell, No. 5837485bert
; APPLICANT: Augustin, Johannes
; APPLICANT: Engelke, Gernar
; APPLICANT: Rosenstein, Ralf
; APPLICANT: Kaletta, Cortina
; APPLICANT: Klein, Cora
; APPLICANT: Wieland, Bernd
; APPLICANT: Kupke, Thomas
; APPLICANT: Jung, G nther
; APPLICANT: Kellner, Roland
; TITLE OF INVENTION: Biosynthetic Process For The Preparation
; TITLE OF INVENTION: Of Chemical Compounds
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/392,625
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/876,791

```
; FILING DATE: 30-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0652.0980002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-392-625-22

Query Match          7.5%; Score 149.5; DB 2; Length 181;
Best Local Similarity 30.9%; Pred. No. 1.1e-06;
Matches 51; Conservative 31; Mismatches 64; Indels 19; Gaps 7;

Qy 40 LIRHGAEVHAYMSEATKIIHPYAWNLTGNPVTETG-FIEHVELAGEHENKADLLV 98
Db 25 LKQHFDEVNLFSSSKNFINDVLKLCFN-LYDEIKDPLNHNINVENHE----YILV 79
Qy 99 CPATANTISKIACGIDDPVTVTAPPHIPIMTAPAMHETMYRHPIVRENIERLKLK 158
Db 80 LPASANTINKANGICDNLTLVCLTGQ--KLFFPNMNMGNPFLQKNIDLLKND 137
Qy 159 VEFIGP-----RIEGRK-----VASIDEIVYRVIKLHKHTLE 193
Db 138 VKVYSPDMNKSFEISSGRYKKNITMPNIENVLNLFVNN-EKRPLD 181

RESULT
US-08-466-961A-22
; Sequence 22, Application US/08466961A
; Patent No. 5843709
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
; APPLICANT: Gtz, Friedrich
; APPLICANT: Schnell, No. 5843709bert
; APPLICANT: Augustin, Johannes
; APPLICANT: Engelke, Gernar
; APPLICANT: Rosenstein, Ralf
; APPLICANT: Kalletta, Cortina
; APPLICANT: Klein, Cora
; APPLICANT: Wieland, Bernd
; APPLICANT: Kupke, Thomas
; APPLICANT: Jung, G nther
; APPLICANT: Kellner, Roland
; TITLE OF INVENTION: Biosynthetic Process for the Preparation of
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466.961A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/392,625
; FILING DATE: 22-FEB-1995
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 07/876,791
; FILING DATE: 30-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/784,234
; FILING DATE: 31-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0652.0980004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2600
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-466-961A-22

Query Match          7.5%; Score 149.5; DB 2; Length 181;
Best Local Similarity 30.9%; Pred. No. 1.1e-06;
Matches 51; Conservative 31; Mismatches 64; Indels 19; Gaps 7;

Qy 40 LIRHGAEVHAYMSEATKIIHPYAWNLTGNPVTETG-FIEHVELAGEHENKADLLV 98
Db 25 LKQHFDEVNLFSSSKNFINTVLKLCFN-LYDEIKDPLNHNINVENHE----YILV 79
Qy 99 CPATANTISKIACGIDDPVTVTAPPHIPIMTAPAMHETMYRHPIVRENIERLKLK 158
Db 80 LPASANTINKANGICDNLTLVCLTGQ--KLFFPNMNMGNPFLQKNIDLLKND 137
Qy 159 VEFIGP-----RIEGRK-----VASIDEIVYRVIKLHKHTLE 193
Db 138 VKVYSPDMNKSFEISSGRYKKNITMPNIENVLNLFVNN-EKRPLD 181

RESULT
US-08-822-774-46
; Sequence 46, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8844
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
```

TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-822-774-46

Query Match 6.8%; Score 135; DB 4; Length 31;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLHHVKLIYATKSRKLVGKIVXXPGSIAA 31
Db 1 MLHHVKLIYATKSRKLVGKIVXXPGSIAA 31

RESULT 8
US-08-822-774-6
Sequence 6, Application US/08822774
Patent No. 6183997
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
TITLE OF INVENTION: and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
ADDRESSEE: Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822.774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-822-774-6

Query Match 5.8%; Score 116; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 353 ADLVVGNTEAFGSEENOVVLGR 376
Db 1 ADLVVGNTEAFGSEENOVVLGR 24

RESULT 9
US-08-822-774-48

Sequence 48, Application US/08822774
Patent No. 6183997
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
TITLE OF INVENTION: and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
ADDRESSEE: Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822.774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 24 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-822-774-48

Query Match 5.8%; Score 116; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 353 ADLVVGNTEAFGSEENOVVLGR 376
Db 1 ADLVVGNTEAFGSEENOVVLGR 24

RESULT 10
US-08-460-309-4
Sequence 4, Application US/08460309
Patent No. 5837496
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/460.309
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/125,077
;; FILING DATE: 22-SEP-1993
;; APPLICATION NUMBER: US PCT/US 94/10730
;; FILING DATE: 21-SEP-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/472,319
;; FILING DATE: 30-JAN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/919,951
;; FILING DATE: 27-JUL-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Campbell, Cathryn A.
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: P-LA 9721
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-8949
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3111 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; US-08-460-309-4

Query Match 5.3%; Score 105; DB 2; Length 3111;
Best Local Similarity 23.3%; Pred. No. 1.4;
Matches 74; Conservative 48; Mismatches 109; Indels 86; Gaps 17;

QY 135 PAMHETMYRHPVIRENI-ERLKLGVFEIGP-----RIEEGRAK--VASIDEIVYRV 183
DB 1604 PAPIKMLYG-----LENMTOELKHL-----LSPQAPERLIQLAEGNLTIVTEMNELLTRA 1655

QY 184 IKLHKHTLEGRVLYTAGATREYIDPIRFITNASS-GKMGVALAEADF--RGAVTLIR 240
DB 1656 TK-----VTADGEQTGDAERTNTRAKSLGEFIKELARDAEAVNEKAIKLNE 1702

QY 241 TKGSVKAFRIKIKLVETVEEMLSAIENELRSKY-----DVIIMAAVSDFRPKI 292
DB 1703 TLGT----RDEAFERNLEGLQKEIDQMIKELRRKNLETKQETADELVAAEALLKKVKKL 1758

QY 293 KAEGKIKS---GRSITIELVPXNPKI-----IDRIKEIQPNVFLVGFKAETSKEK 339
DB 1759 FGESGNEEMEKDLREKLADYKNKVDADWLLREATDKIREAN-RLFAVNQKNNTALEK 1817

QY 340 ---LIEEGKQRIERAKADLVVGNLT-----EAFGSENOVVLIGRDTFKELPKM----- 385
DB 1818 KKEAVESGKQRIENT---LKEGNDILDEANRLADEINSIIDYVEDIQTKLPPMSEELNDK 1874

QY 386 -----KKRELAERI 394
DB 1875 IDDLSQEIKDKRLAEKV 1891

RESULT 11
us-08-125-077-4
; Sequence 4, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof

;; NUMBER OF SEQUENCES: 23
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Campbell and Flores
;; STREET: 4370 La Jolla Village Drive, Suite 700
;; CITY: San Diego
;; STATE: California
;; COUNTRY: USA
;; ZIP: 92122
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/125,077
;; FILING DATE: 22-SEP-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US PCT/US 94/10730
;; FILING DATE: 21-SEP-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/472,319
;; FILING DATE: 30-JAN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/919,951
;; FILING DATE: 27-JUL-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Campbell, Cathryn A.
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: P-LA 9721
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-8949
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3111 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; US-08-125-077-4

Query Match 5.3%; Score 105; DB 2; Length 3111;
Best Local Similarity 23.3%; Pred. No. 1.4;
Matches 74; Conservative 48; Mismatches 109; Indels 86; Gaps 17;

QY 135 PAMHETMYRHPVIRENI-ERLKLGVFEIGP-----RIEEGRAK--VASIDEIVYRV 183
DB 1604 PAPIKMLYG-----LENMTOELKHL-----LSPQAPERLIQLAEGNLTIVTEMNELLTRA 1655

QY 184 IKLHKHTLEGRVLYTAGATREYIDPIRFITNASS-GKMGVALAEADF--RGAVTLIR 240
DB 1656 TK-----VTADGEQTGDAERTNTRAKSLGEFIKELARDAEAVNEKAIKLNE 1702

QY 241 TKGSVKAFRIKIKLVETVEEMLSAIENELRSKY-----DVIIMAAVSDFRPKI 292
DB 1703 TLGT----RDEAFERNLEGLQKEIDQMIKELRRKNLETKQETADELVAAEALLKKVKKL 1758

QY 293 KAEGKIKS---GRSITIELVPXNPKI-----IDRIKEIQPNVFLVGFKAETSKEK 339
DB 1759 FGESGNEEMEKDLREKLADYKNKVDADWLLREATDKIREAN-RLFAVNQKNNTALEK 1817

QY 340 ---LIEEGKQRIERAKADLVVGNLT-----EAFGSENOVVLIGRDTFKELPKM----- 385
DB 1818 KKEAVESGKQRIENT---LKEGNDILDEANRLADEINSIIDYVEDIQTKLPPMSEELNDK 1874

QY 386 -----KKRELAERI 394
DB 1875 IDDLSQEIKDKRLAEKV 1891

RESULT 12
US-08-822-774-3
; Sequence 3, Application US/08822774

; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-822-774-3

Query Match 5.2%; Score 104; DB 4; Length 35;
Best Local Similarity 70.0%; Pred. No. 0.0021;
Matches 21; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
Qy 1 MLHHVKLIYATKSKLYGKKIVXXPGSIA 30
Db 2 LLHHVKLIYATKXXLVGKXIVLAIPGXXA 31

RESULT 13
US-08-317-223-3
; Sequence 3, Application US/08317223
; Patent No. 5585267
; GENERAL INFORMATION:
; APPLICANT: Jones, Jonathan C.R.
; APPLICANT: Quaranta, Vito
; APPLICANT: Tamura, Richard
; TITLE OF INVENTION: CELLULAR ATTACHMENT TO TRANS-EPITHELIAL
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,223
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,727
; FILING DATE: 05-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/151,134
; FILING DATE: 12-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/152,460
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: DESMOS.002CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; IMMEDIATE SOURCE:
; CLONE: merosin
; US-08-317-223-3

Query Match 5.1%; Score 102.5; DB 1; Length 337;
Best Local Similarity 23.4%; Pred. No. 0.085;
Matches 65; Conservative 42; Mismatches 102; Indels 69; Gaps 13;
Qy 165 RIEGRK--VASIDEIVYRVIKLHKLTLEGKRVLTAGATREYIDPIRITWASS-GK 221
Db 17 QLAEGNLTLVTEMNELLTRATK-----VTADGEOTGQDAERTNTRAKSLGE 63
Qy 222 MGVALAEADF--RGVTLIFTGSKVAFIRKIKLVETVEEMLSAIENELRSKKY--- 276
Db 64 FIKELARDAEAVNEKAIKLNETLGT----RDEAFERNLEGQKEIDQMIKELRKNLEQ 119
Qy 277 -----DVVIMAAVSDFRPKIKAEKIKS---GRSITIELVXPXPKI-----IDR 318
Db 120 KEIAEDELVAEALKKVKLFGESGENEEMKDLREKLADYKNKYDDAWDLLREATDK 179
Qy 319 IKEIQPNVFLVGFKAETSKK--LIEEGKQIERAKADLVVGNLT-----EAFGSEENQV 371
Db 180 IREAN-RLFAVNQKNMTALEKKKEAVESGKQIEN---LKEGNDILDEANRLADEINSI 235
Qy 372 VLIGRDFTKELPKM-----KKRELAERI 394
Db 236 IDYVEDIQTKLPPMSEELNDKIDLSQEIKDRKLAEKV 273

RESULT 14
US-09-059-849A-3
; Sequence 3, Application US/09059849A
; Patent No. 6034068
; GENERAL INFORMATION:
; APPLICANT: Craig Halberstadt
; TITLE OF INVENTION: CELLULAR ATTACHMENT TO LAMININ 5-COATED
; TITLE OF INVENTION: TRANS-EPITHELIAL APPLIANCES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor

```

; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,849A
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: DESMOS.015A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; IMMEDIATE SOURCE:
; CLONE: merosin
;
; US-09-059-849A-3

```

```

Query Match          5.1%; Score 102.5; DB 3; Length 337;
Best Local Similarity 23.4%; Pred. No. 0.085;
Matches 65; Conservative 42; Mismatches 102; Indels 69; Gaps 13;

QY 165 RIEGRK--VASIDEIVYRVIKLHKHKTLEGKRVLVTAGATREYIDPIRFTINASS-GK 221
DB 17 QLAEGNLTIVTEMNELLTRATK-----VTADGQTQDAERTNTRAKSLGE 63
QY 222 MGVALAEADF--RGAVTLIRTKGSKVAFRIKIKLVETVEEMLSAIENELRSKKY--- 276
DB 64 FIKELARDAEAVNEKAIKLNETLGT-----RDEAFERNLEGLQKEIDQMIKELRRKLE 119
QY 277 -----DVVIMAAVSDFRPKIKAEKGIKS---GRSITIELVPXNPKI-----IDR 318
DB 120 KETADELVAEALLKKVKLGESRGNEEMKDLREKLADYKNKVDDADWDLLEATDK 179
QY 319 IKEIQPNVFLVGFKAETSKK---LIEEGKQIERAKADLVVGNLT-----EAFGSEENOV 371
DB 180 IREAN-RLFAVNQKNWTALEKKKKEAVESGKQIENT---LKEGNDILDEANRLADEINSI 235
QY 372 VLIGRDFTKELPKM-----KRELAERI 394
DB 236 IDYVEDIQTKLPPMSEELNDKIDLSQEIQRKLAEKV 273

```

```

RESULT 15
US-09-213-632-3
; Sequence 3, Application US/092113632
; Patent No. 6110711
; GENERAL INFORMATION:
; APPLICANT: Jones, Jonathan C.R.
; APPLICANT: Quaranta, Vito
; APPLICANT: Tamura, Richard
; TITLE OF INVENTION: CELLULAR ATTACHMENT TO TRANS-EPITHELIAL
; TITLE OF INVENTION: APPLIANCES

```

```

; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/213,632
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/317,223
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/151,134
; FILING DATE: 12-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/152,460
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: DESMOS.002CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; IMMEDIATE SOURCE:
; CLONE: merosin
;
; US-09-213-632-3

```

```

Query Match          5.1%; Score 102.5; DB 3; Length 337;
Best Local Similarity 23.4%; Pred. No. 0.085;
Matches 65; Conservative 42; Mismatches 102; Indels 69; Gaps 13;

QY 165 RIEGRK--VASIDEIVYRVIKLHKHKTLEGKRVLVTAGATREYIDPIRFTINASS-GK 221
DB 17 QLAEGNLTIVTEMNELLTRATK-----VTADGQTQDAERTNTRAKSLGE 63
QY 222 MGVALAEADF--RGAVTLIRTKGSKVAFRIKIKLVETVEEMLSAIENELRSKKY--- 276
DB 64 FIKELARDAEAVNEKAIKLNETLGT-----RDEAFERNLEGLQKEIDQMIKELRRKLE 119
QY 277 -----DVVIMAAVSDFRPKIKAEKGIKS---GRSITIELVPXNPKI-----IDR 318
DB 120 KETADELVAEALLKKVKLGESRGNEEMKDLREKLADYKNKVDDADWDLLEATDK 179
QY 319 IKEIQPNVFLVGFKAETSKK---LIEEGKQIERAKADLVVGNLT-----EAFGSEENOV 371
DB 180 IREAN-RLFAVNQKNWTALEKKKKEAVESGKQIENT---LKEGNDILDEANRLADEINSI 235
QY 372 VLIGRDFTKELPKM-----KRELAERI 394
DB 236 IDYVEDIQTKLPPMSEELNDKIDLSQEIQRKLAEKV 273

```


Search completed: January 31, 2002, 13:07:51
Job time: 43 sec

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:02 ; Search time 140.03 Seconds
(without alignments)
213.180 Million cell updates/sec

Title: US-08-957-709-19
Perfect score: 403
Sequence: 1 MLHHVKLIYATKSRKLVGRK.....MKMKRELAERIWDEIEKXLS 403

Scoring table: OLIGO
Gapop 60.0 , Gapert 60.0

Searched: 522463 seqs, 74073290 residues

Word size: 0
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_ll01.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length DB ID	Description	
1	297	73.7	403 19	AAW72844	Polymerase enhanci
2	28	6.9	31 19	AAW72841	Polymerase enhanci
3	24	6.0	24 19	AAW72863	Polymerase enhanci
4	24	6.0	24 19	AAW72843	Polymerase enhanci
5	17	4.2	17 19	AAW72842	Polymerase enhanci
6	13	3.2	35 19	AAW72860	Polymerase enhanci
7	12	3.0	13 19	AAW72866	Polymerase enhanci
8	11	2.7	15 19	AAW72864	Polymerase enhanci
9	11	2.7	15 19	AAW72865	Polymerase enhanci
10	10	2.5	15 19	AAW72858	Polymerase enhanci
11	8	2.0	8 19	AAW72871	Polymerase enhanci

12	8	2.0	463	22	AAE04541	Aspergillus fumiga
13	8	2.0	481	18	AAW55465	H. pylori ORF 05ae
14	8	2.0	486	18	AAW55242	H. pylori ORF 05ae
15	8	2.0	490	21	AAW04881	Arabidopsis thalia
16	8	2.0	491	21	AAW53291	Arabidopsis thalia
17	8	2.0	505	21	AAW04880	Arabidopsis thalia
18	8	2.0	506	21	AAW53290	Arabidopsis thalia
19	8	2.0	511	21	AAW04879	Arabidopsis thalia
20	8	2.0	511	21	AAW53289	Arabidopsis thalia
21	7	1.7	7	19	AAW72873	Polymerase enhanci
22	7	1.7	17	19	AAW72862	Polymerase enhanci
23	7	1.7	18	22	AAW92176	Signal transductio
24	7	1.7	105	18	AAW28131	Pyruvate dehydrogen
25	7	1.7	115	18	AAW11301	S. pneumoniae prot
26	7	1.7	143	21	AAW36086	Zea mays protein f
27	7	1.7	144	21	AAW36085	Zea mays protein f
28	7	1.7	169	21	AAW36084	Zea mays protein f
29	7	1.7	170	21	AAW35412	Zea mays protein f
30	7	1.7	172	22	AAW14877	Peptide #1311 enco
31	7	1.7	172	22	AAW14880	Peptide #1314 enco
32	7	1.7	172	22	AAW27306	Peptide #1343 enco
33	7	1.7	172	22	AAW27309	Peptide #1346 enco
34	7	1.7	172	22	AAW02602	Peptide #1284 enco
35	7	1.7	172	22	AAW02604	Peptide #1286 enco
36	7	1.7	174	21	AAW11769	Arabidopsis thalia
37	7	1.7	174	21	AAW46241	Arabidopsis thalia
38	7	1.7	181	21	AAW11768	Arabidopsis thalia
39	7	1.7	181	21	AAW46240	Arabidopsis thalia
40	7	1.7	182	22	AAW42118	Human polypeptide
41	7	1.7	182	22	AAW42119	Human polypeptide
42	7	1.7	194	21	AAW82744	DNA replication an
43	7	1.7	196	20	AAW35142	Chlamydia pneumoni
44	7	1.7	203	18	AAW33110	Cell cycle regulat
45	7	1.7	206	21	AAW11767	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
ID	AAW72844 standard; Protein; 403 AA.
AC	AAW72844;
XX	AC
DT	01-MAR-1999 (first entry)
XX	DT
DE	Polymerase enhancing factor P50 component.
XX	DE
KW	Polymerase enhancing factor; PEF; DNA polymerase; PCR;
KW	amplification; sequencing; replication.
OS	Pyrococcus furiosus strain DSM 3638.
XX	OS
FH	Key Location/Qualifiers
FT	Misc-difference 23 /note= "encoded by NNN"
FT	Misc-difference 24 /note= "encoded by NNN"
FT	Misc-difference 25 /note= "encoded by NNN"
FT	Misc-difference 310 /note= "encoded by CCN"
FT	Misc-difference 311 /note= "encoded by NNN"
FT	Peptide 1..31
FT	Peptide /note= "N-terminal peptide used to generate primers"
FT	Peptide 275..291
FT	Peptide /note= "internal peptide used to generate primers"
FT	Peptide 353..376
FT	Peptide /note= "internal peptide used to generate primers"

1 adlvvntleafgseenavljgr 24

```

RESULT 5
AAW72842
ID AAW72842 standard; Peptide: 17 AA.
XX AC
XX AAW72842;
XX DT
XX DE
XX 01-MAR-1999 (first entry)
XX DE
XX Polymerase enhancing factor P50 component internal peptide.
XX KW
XX Polymerase enhancing factor; PEF; DNA polymerase; PCR;
XX amplification; sequencing; replication.
XX OS
XX Pyrococcus furiosus strain DSM 3638.
XX PN
XX Key Location/Qualifiers
XX Misc-difference 1 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 14 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 16 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 21 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 29 /note= "any amino acid"
FT
XX Misc-difference 30 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 32..35 /note= "these residues may each be deleted or
FT substituted by any amino acid but are
FT tentatively assigned as indicated"
XX WO9842860-A1.
XX 01-OCT-1998.
XX 20-MAR-1998; 98WO-US05497.
XX 24-OCT-1997; 97US-0957709.
XX 21-MAR-1997; 97US-0822774.
XX (STRA-) STRATAGENE.
XX Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
XX Polymerase enhancing factor proteins, extracts and complexes -
FT improve the polymerisation activity of nucleic acid polymerases, for
FT use in amplification, sequencing and replication
XX Claim 11; Page 91; 161pp; English.
XX This peptide comprises an internal peptide of the P50 component
CC of the polymerase enhancing factor (PEF) of Pyrococcus furiosus
CC DSM 3638. It is obtained from the translated sequence (see AAW72844)
CC of isolated P50 DNA (see AAV63859) and corresponds to a chemically
CC determined N-terminal peptide (see AAW72862). PEF, the predominant
CC components of which are P50 and P45 (see AAW72847) proteins, enhances
CC the activity of P. furiosus DNA polymerase, thereby providing
CC replication products of greater length and purity. The invention
CC provides novel extracts, proteins and complexes that improve the
CC polymerisation activity of nucleic acid polymerases. These
CC complexes may include proteins including the P50 internal peptide.
CC Also included are methods for identifying compositions with
CC polymerase enhancing activity, for purifying and using these
CC compositions, and specific extracts, proteins and complexes that
CC function to enhance polymerase activity. Nucleic acid polymerase
CC reactions can be enhanced (claimed) by mixing a nucleic acid
CC template, at least 1 polymerase and a composition having polymerase
CC enhancing activity. Kits are provided for replicating nucleic
CC acids. The kits can be used in site-directed mutagenesis, nucleic
CC acid sequencing or amplification (preferably PCR or RT-PCR).
XX SQ Sequence 17 AA;

Query Match 4.28; Score 17; DB 19; Length 17;
Best Local Similarity 100.0%; Pred. NO. 5e-10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 275 KYDVVIMAAVSDFRPK 291
Db 1 kydvvimaaavsdfrpk 17
|||||
|||||

RESULT 6
AAW72860
ID AAW72860 standard; Peptide: 35 AA.
XX AC
XX AAW72860;
XX DT
XX DE
XX 01-MAR-1999 (first entry)
XX DE
XX Polymerase enhancing factor P50 component N-terminal peptide.
XX KW
XX Polymerase enhancing factor; PEF; DNA polymerase; PCR;
XX amplification; sequencing; replication.
XX OS
XX Pyrococcus furiosus strain DSM 3638.
XX PN
XX Key Location/Qualifiers
XX Misc-difference 1 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 14 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 16 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 21 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 29 /note= "any amino acid"
FT
XX Misc-difference 30 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX Misc-difference 32..35 /note= "these residues may each be deleted or
FT substituted by any amino acid but are
FT tentatively assigned as indicated"
XX WO9842860-A1.
XX 01-OCT-1998.
XX 20-MAR-1998; 98WO-US05497.
XX 24-OCT-1997; 97US-0957709.
XX 21-MAR-1997; 97US-0822774.
XX (STRA-) STRATAGENE.
XX Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
XX Polymerase enhancing factor proteins, extracts and complexes -
FT improve the polymerisation activity of nucleic acid polymerases, for
FT use in amplification, sequencing and replication
XX Example 5; Page 32; 161pp; English.
XX This is an N-terminal peptide of the P50 component of the
CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638.
CC It was obtained by N-terminal sequencing of P50 isolated from a
CC PEF complex. PEF, the predominant components of which are P50 (see
CC AAW72844) and P45 (see AAW72847), enhances the activity of P. furiosus
CC DNA polymerase. The invention provides novel extracts, proteins
CC and complexes that improve the polymerisation activity of nucleic
CC acid polymerases. These can be used to improve nucleic acid
CC replication, polymerisation and amplification (especially in PCR or
CC RT-PCR).
XX SQ Sequence 35 AA;

```

```

Query Match      3.2%; Score 13; DB 19; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATKSR 14
   | | | | | | | | | |
DB 3 lhhvkliyatksr 15

RESULT 7
AAW72866
ID AAW72866 standard; Peptide; 13 AA.
XX
AC AAW72866;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P50 component N-terminal peptide.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR;
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
PN WO9842860-Al.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
XX
PR 21-MAR-1997; 97US-0822774.
XX
PA (STRA-) STRATAGENE.
XX
PI Hansen CJ, Hogrefe H;
XX
DR WPI; 1998-542284/46.
XX
PT Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication
XX
PS Example 5; Page 33; 16lpp; English.
XX
CC This is an N-terminal peptide of the P50 component of the
CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638.
CC It was obtained by N-terminal sequencing of P50 isolated from a
CC PEF complex. PEF, the predominant components of which are P50 (see
CC AAW72844) and P45 (see AAW72847), enhances the activity of P. furiosus
CC DNA polymerase. The invention provides novel extracts, proteins
CC and complexes that improve the polymerisation activity of nucleic
CC acid polymerases. These can be used to improve nucleic acid
CC replication, polymerisation and amplification (especially in PCR or
CC RT-PCR).
XX
SQ Sequence 13 AA;

Query Match      3.0%; Score 12; DB 19; Length 13;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHHVKLIYATK 12
   | | | | | | | | | |
DB 1 mlhhvkliyatck 12

RESULT 8
AAW72864
ID AAW72864 standard; Peptide; 15 AA.
XX

```

```

AC AAW72864;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor 150 kDa protein N-terminal peptide.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR;
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
PH Key Location/Qualifiers
FT Misc-difference 1 /label= Gly, Ala, Met
FT FT
FT Misc-difference 11 /note= "this residue may be deleted or substituted
FT by any amino acid but is tentatively
FT assigned as indicated"
XX
PN WO9842860-Al.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
XX
PR 21-MAR-1997; 97US-0822774.
XX
PA (STRA-) STRATAGENE.
XX
PI Hansen CJ, Hogrefe H;
XX
DR WPI; 1998-542284/46.
XX
PT Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication
XX
PS Example 5; Page 33; 16lpp; English.
XX
CC This is an N-terminal peptide of a 150 kDa component of the
CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638.
CC PEF, the predominant components of which are P50 (see AAW72844) and
CC P45 (see AAW72847), enhances the activity of P. furiosus DNA
CC polymerase. The invention provides novel extracts, proteins and
CC complexes that improve the polymerisation activity of nucleic acid
CC polymerases. These can be used to improve nucleic acid replication,
CC polymerisation and amplification (especially in PCR or RT-PCR).
XX
SQ Sequence 15 AA;

Query Match      2.7%; Score 11; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATK 12
   | | | | | | | | | |
DB 2 lhhvkliyatck 12

RESULT 9
AAW72865
ID AAW72865 standard; Peptide; 15 AA.
XX
AC AAW72865;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor 100 kDa protein N-terminal peptide.
XX
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR;
XX

```

```

XX OS Pyrococcus furiosus strain DSM 3638.
XX PH Key Location/Qualifiers
XX FT Misc-difference 1 /label= Gly, Ala, Met
XX FT Misc-difference 13 /label= Lys, Leu
XX FT /note= "these residues may each be deleted or
XX FT substituted by any amino acid but are
XX FT assigned as indicated"
XX PN WO9842860-A1.
XX PD 01-OCT-1998.
XX PF 20-MAR-1998; 98WO-US05497.
XX PR 24-OCT-1997; 97US-0957709.
XX PR 21-MAR-1997; 97US-0822774.
XX PA (STRA-) STRATAGENE.
XX PI Hansen CJ, Hogrefe H;
XX DR WPI; 1998-542284/46.
XX PT Polymerase enhancing factor proteins, extracts and complexes -
XX FT improve the polymerisation activity of nucleic acid polymerases, for
XX FT use in amplification, sequencing and replication
XX PS Example 5; Page 33; 161pp; English.
XX CC This is an N-terminal peptide of a 100 kDa component of the
XX CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638.
XX CC P45, the predominant components of which are P50 (see AAW72844) and
XX CC P46 (see AAW72847), enhances the activity of P. furiosus DNA
XX CC polymerase. The invention provides novel extracts, proteins and
XX CC complexes that improve the polymerisation activity of nucleic acid
XX CC polymerases. These can be used to improve nucleic acid replication,
XX CC polymerisation and amplification (especially in PCR or RT-PCR).
XX SQ Sequence 15 AA;

Query Match 2.7%; Score 11; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYATK 12
Db 2 lhhvkliyat 12

RESULT 10
AAW72858
ID AAW72858 standard; Peptide; 15 AA.
AC AAW72858;
DT 01-MAR-1999 (first entry)
DE Polymerase enhancing factor P50 component N-terminal peptide.
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR;
KW amplification; sequencing; replication.
OS Pyrococcus furiosus strain DSM 3638.
XX PH Key Location/Qualifiers
XX FT Misc-difference 1 /note= "any amino acid"
XX FT Misc-difference 2 /note= "any amino acid"
XX PN WO9842860-A1.

```

```

XX OS Pyrococcus furiosus strain DSM 3638.
XX PH Key Location/Qualifiers
XX FT Misc-difference 13 /note= "this residue may be deleted or substituted
XX FT by any amino acid but is tentatively
XX FT assigned as indicated"
XX FT Misc-difference 14 /note= "any amino acid"
XX FT Misc-difference 15 /note= "any amino acid"
XX FT /note= "this residue may be deleted or substituted
XX FT by any amino acid but is tentatively
XX FT assigned as indicated"
XX PN WO9842860-A1.
XX PD 01-OCT-1998.
XX PF 20-MAR-1998; 98WO-US05497.
XX PR 24-OCT-1997; 97US-0957709.
XX PR 21-MAR-1997; 97US-0822774.
XX PA (STRA-) STRATAGENE.
XX PI Hansen CJ, Hogrefe H;
XX DR WPI; 1998-542284/46.
XX PT Polymerase enhancing factor proteins, extracts and complexes -
XX FT improve the polymerisation activity of nucleic acid polymerases, for
XX FT use in amplification, sequencing and replication
XX PS Example 5; Page 32; 161pp; English.
XX CC This is an N-terminal peptide of the P50 component of the
XX CC polymerase enhancing factor (PEF) of Pyrococcus furiosus DSM 3638.
XX CC It was obtained by N-terminal sequencing of P50 isolated from a
XX CC PEF complex. PEF, the predominant component of which are P50 (see
XX CC AAW72844) and P45 (see AAW72847), enhances the activity of P. furiosus
XX CC DNA polymerase. The invention provides novel extracts, proteins
XX CC and complexes that improve the polymerisation activity of nucleic
XX CC acid polymerases. These can be used to improve nucleic acid
XX CC replication, polymerisation and amplification (especially in PCR or
XX CC RT-PCR).
XX SQ Sequence 15 AA;

Query Match 2.5%; Score 10; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.006;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LHHVKLIYAT 11
Db 3 lhhvkliyat 12

RESULT 11
AAW72871
ID AAW72871 standard; Peptide; 8 AA.
AC AAW72871;
DT 01-MAR-1999 (first entry)
DE Polymerase enhancing factor P50 component N-terminal peptide.
KW Polymerase enhancing factor; PEF; DNA polymerase; PCR;
KW amplification; sequencing; replication.
OS Pyrococcus furiosus strain DSM 3638.
XX PN WO9842860-A1.

```


PD 01-OCT-1998.
 XX
 PF 20-MAR-1998; 98WO-US05497.
 XX
 PR 24-OCT-1997; 97US-0957709.
 PR 21-MAR-1997; 97US-0822774.
 XX
 PA (STRA-) STRATAGENE.
 XX
 PI Hansen CJ, Hogrefe H;
 XX WPI; 1998-542284/46.
 DR
 XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 PT use in amplification, sequencing and replication
 XX
 PS Example 6; Page 34; 161pp; English.
 XX
 CC This peptide is derived from an N-terminal peptide (see AAW72858) of
 CC the P50 component of the polymerase enhancing factor (PEF) of
 CC *Pyrococcus furiosus* DSM 3638. A degenerate primer (see AAV63871)
 CC based on the peptide has been used to amplify DNA (see AAV63859)
 CC coding for P50 protein (see AAW72844) from *P. furiosus* genomic DNA.
 CC PEF enhances the activity of *P. furiosus* DNA polymerase. The
 CC invention provides novel extracts, proteins and complexes that
 CC improve the polymerisation activity of nucleic acid polymerases.
 CC These can be used to improve nucleic acid replication,
 CC polymerisation and amplification (especially in PCR or RT-PCR).
 XX
 SQ Sequence 8 AA;

Query Match 2.0%; Score 8; DB 19; Length 8;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 HHVKLIYA 10
 |||||
 Db 1 hhvkliya 8

RESULT 12
 AAE04541
 ID AAE04541 standard; Protein; 463 AA.
 XX
 AC AAE04541;
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Aspergillus fumigatus N-myristoyl transferase (NMT).
 XX
 KW N-myristoyl transferase; NMT; identification; anti-fungal agent;
 KW cotranslational modification.
 XX
 OS Aspergillus fumigatus.
 XX
 PN US6251596-B1.
 XX
 PD 26-JUN-2001.
 XX
 PF 30-SEP-1998; 98US-0163444.
 XX
 PR 30-SEP-1998; 98US-0163444.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Cook WJ, Bulawa CE;
 XX
 DR WPI; 2001-431958/46.
 DR N-PSDB; AAD09840.
 XX
 PT New isolated nucleic acid molecules, which encode *Aspergillus fumigatus*

PT N-myristoyl transferase, useful for identifying anti-fungal agents -
 XX
 PS Claim 1; Fig 1; 23pp; English.
 XX
 CC The present invention relates to *Aspergillus fumigatus* N-myristoyl
 CC transferase (NMT) gene and its use in identifying anti-fungal agents.
 CC NMT is responsible for cotranslational modification of a variety of
 CC fungal proteins. NMT catalyses the attachment of a 14-carbon saturated
 CC fatty acid to the N-terminal glycine residue of cellular proteins. This
 CC modification is thought to be irreversible and essential for the full
 CC biological activity of myristoylated proteins. The present sequence is
 CC *Aspergillus fumigatus* N-myristoyl transferase (NMT).
 XX
 SQ Sequence 463 AA;

Query Match 2.0%; Score 8; DB 22; Length 463;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 ATKSRKLV 17
 |||||
 Db 154 atksrkvl 161

RESULT 13
 AAW55465
 ID AAW55465 standard; Protein; 481 AA.
 XX
 AC AAW55465;
 XX
 DT 24-JUN-1998 (first entry)
 XX
 DE H. pylori ORF 05ae30220_24882812_c3_103 secreted protein.

KW Cytoplasmic; vaccine; prevention; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 XX
 OS *Helicobacter pylori*.
 XX
 PN WO9737044-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-US05223.
 XX
 PR 06-DEC-1996; 96US-0761318.
 PR 29-MAR-1996; 96US-0625811.
 PR 02-APR-1996; 96US-0758731.
 PR 25-OCT-1996; 96US-0736905.
 PR 28-OCT-1996; 96US-0738859.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Alm RA, Smith D;
 XX
 DR WPI; 1997-503122/46.
 DR N-PSDB; AAV24874.
 XX
 PT *Helicobacter pylori* nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent *H. pylori*
 PT infection and for diagnosis of *H. pylori* infection
 XX
 PS Claims 14,94; Page 671-672; 1145pp; English.
 XX
 CC This sequence is a *H. pylori* secreted protein.
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*
 CC infection or to identify *H. pylori* polypeptide binding compounds, the
 CC useful as potential *H. pylori* life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC *H. pylori* in a sample and the diagnosis of *H. pylori* infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and

CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genetic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by "randomly shearing the bacterial DNA. The sequences were analysed
CC for ORFs at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
XX hosts.

XX Sequence 481 AA;
SQ
Query Match 2.0%; Score 8; DB 18; Length 481;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 336 SKEKLEIE 343
Db 281 skeklee 288
|||||||

RESULT 14
AAW55242
ID AAW55242 standard; Protein; 486 AA.
AC AAW55242;

DE 26-JUN-1998 (first entry)
H. pylori ORF 05ae20220orf119 protein.

CC Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
CC identification; binding compound; bacteria; life cycle; activator;
CC inhibition; duodenal ulcer disease; chronic gastritis; diagnosis;
CC bacterium.

XX Helicobacter pylori.
XX WO9737044-A1.
XX 09-OCT-1997.

XX 27-MAR-1997; 97WO-US05223.
XX 06-DEC-1996; 96US-0761318.
XX 29-MAR-1996; 96US-0625811.
XX 02-APR-1996; 96US-0758731.
XX 25-OCT-1996; 96US-0736905.
XX 28-OCT-1996; 96US-0738859.

XX (ASTR) ASTRA AB.
XX Alm RA, Smith D;
XX WPI; 1997-503122/46.
XX N-PSDB; AAV24651.

XX Helicobacter pylori nucleic acid sequences and encoded
XX polypeptide(s) - useful in vaccines to treat or prevent H. pylori
XX infection and for diagnosis of H. pylori infection

XX Claim 14; Pages 485-486; 1145pp; English.

XX This sequence is a Helicobacter pylori protein of unspecified
XX function.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds,
XX useful as potential H. pylori life cycle activators or inhibitors.

CC The DNA and probes derived from it may be used for the
CC identification of H. pylori in a sample, and the diagnosis of
CC H. pylori infection. Nucleic acid sequences complementary to the
CC DNA act as antisense sequences, and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can
CC be used in immunoassays to evaluate the abundance and distribution
CC of H. pylori-specific antigens. The genomic sequence of H. pylori
CC (ATCC 55679) was determined from overlapping contigs generated by
CC mechanically shearing the bacterial DNA. The predicted coding
CC regions for ORF of at least 180 nucleotides, and the predicted
CC coding regions defined by computer evaluation, and the predicted
CC H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having
CC identified and determined the sequences of interest, particular
CC regions can be isolated from H. pylori by PCR amplification for
CC recombinant polypeptide production, e.g. in E. coli hosts.

SQ Sequence 486 AA;

Query Match 2.0%; Score 8; DB 18; Length 486;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 336 SKEKLEIE 343
Db 286 skeklee 293
|||||||

RESULT 15
AAG04881
ID AAG04881 standard; Protein; 490 AA.

AC AAG04881;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1075.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 23-MAR-1999; 99US-0123548.
XX 25-MAR-1999; 99US-0125788.
XX 29-MAR-1999; 99US-0126264.
XX 01-APR-1999; 99US-0126785.
XX 06-APR-1999; 99US-0127462.
XX 08-APR-1999; 99US-0128234.
XX 16-APR-1999; 99US-0128714.
XX 19-APR-1999; 99US-0129845.
XX 21-APR-1999; 99US-0130077.
XX 23-APR-1999; 99US-0130519.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 31-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.

```

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      2.0%; Score 8; DB 21; Length 490;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 149 ENTERLKK 156
Db 464 enterlkk 471

```

Search completed: January 31, 2002, 13:18:03
 Job time: 164 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:02 ; Search time 78.64 seconds
(without alignments)
390.366 Million cell updates/sec

Title: US-08-957-709-19
Perfect score: 403
Sequence: 1 MLHHVKLIYATKSRKLVGKK.....KMKRELAERIWEIEKXLS 403

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	72	17.9	401	2 G71018	pantothenate metab
2	60	14.9	401	2 E75114	dna/pantothenate m
3	12	3.0	394	2 A72223	pantothenate metab
4	11	2.7	403	2 A64414	pantothenate metab
5	10	2.5	394	2 D81819	hypothetical prote
6	10	2.5	394	2 A81058	DNA/pantothenate m
7	9	2.2	404	2 D69455	pantothenate metab
8	8	2.0	127	2 T23443	hypothetical prote
9	8	2.0	210	2 S62355	high mobility grou
10	8	2.0	391	2 E71125	probable aromatic-
11	8	2.0	396	2 T36548	probable transamin
12	8	2.0	480	1 TWWTGC	phosphoglycerate k
13	8	2.0	481	2 B64585	cag pathogenicity
14	8	2.0	481	2 D71927	cag island protein
15	8	2.0	794	2 T36972	probable membrane
16	8	2.0	1261	2 S75130	sensory transducti
17	7	1.7	36	2 T03348	gene e19 protein -
18	7	1.7	60	2 G69115	hypothetical prote
19	7	1.7	91	2 F69252	hypothetical prote
20	7	1.7	101	2 B85774	hypothetical prote
21	7	1.7	111	2 T70338	methionine--trna l
22	7	1.7	124	2 C70537	hypothetical prote
23	7	1.7	130	2 F75075	hypothetical prote
24	7	1.7	138	2 C64382	hypothetical prote
25	7	1.7	140	2 E75056	probable translati
26	7	1.7	140	2 D71104	probable translati
27	7	1.7	144	2 A96580	hypothetical prote
28	7	1.7	145	2 E71723	ribosomal protein
29	7	1.7	160	2 T10278	protein tyrosine p

30	7	1.7	170	2 T43779	ribosomal protein
31	7	1.7	173	2 G82818	disulfide bond for
32	7	1.7	186	2 B61611	nonhistone chromos
33	7	1.7	191	2 D86555	CLP proteinase [im
34	7	1.7	191	2 G72067	endopeptidase Clp
35	7	1.7	196	2 B81672	endopeptidase Clp
36	7	1.7	203	2 T38480	hypothetical prote
37	7	1.7	209	2 A69864	conserved hypotet
38	7	1.7	236	2 A37830	luxG protein - Vib
39	7	1.7	241	2 D72304	phosphoribosylform
40	7	1.7	252	2 D64307	5'-methylthioadeno
41	7	1.7	255	2 T36778	probable enoyl-(ac
42	7	1.7	255	2 T50916	hypothetical prote
43	7	1.7	268	2 G81257	probable hemein up
44	7	1.7	270	2 T48229	hypothetical prote
45	7	1.7	274	2 D83425	conserved hypotet

ALIGNMENTS

RESULT 1

G71018

pantothenate metabolism flavoprotein dfp homolog PH1444 - Pyrococcus horikoshii
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: G71018

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: G71018

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-401 <KAW>

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30551.1; PID:g3257868

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH1444

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 17.9%; Score 72; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 1e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 TGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIAGIDTPVTVVTTAF 127

Db 69 TGNPVITEITGFIHVELAGHENKADLILVCPATANTISKIAGIDTPVTVVTTAF 128

Qy 128 HIPIMIAPAMHE 139

Db 129 HIPIMIAPAMHE 140

RESULT 2

E75114

dna/pantothenate metabolism flavoprotein (dfp) PAB1897 - Pyrococcus abyssi (strain Or
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: E75114

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: E75114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-401 <KAW>

A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49630.1; PID:g545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1897

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 14.9%; Score 60; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 6.7e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 80 IEHVELAGEHENKADILVCPATANTISKIACGGIDDTPTVTVTAFPHIPIMIAPAMHE 139

Db 81 IEHVELAGEHENKADILVCPATANTISKIACGGIDDTPTVTVTAFPHIPIMIAPAMHE 140

RESULT 3

A72223

N:Alternate names: pantothenate metabolism flavoprotein dfp homolog TM1687 - Thermotoga maritima (strain MS

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: A72223

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,

Carrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; MUID:99287316

A:Accession: A72223

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <ARN>

A:Cross-references: GB:AF001809; GB:AB000512; NID:g4982257; PIDN:AAD36754.1; PID:g498226

A:Experimental source: strain MSB

C:Genetics:

A:Gene: TM1687

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 3.0%; Score 12; DB 2; Length 394;

Best Local Similarity 100.0%; Pred. No. 0.0014;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 212 RFITNASSGKMG 223

Db 200 RFITNASSGKMG 211

RESULT 4

A6414

N:Alternate names: pantothenate metabolism flavoprotein MJ0913 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: A6414

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,

Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;

rsen, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A:Reference number: A64300; MUID:96337999

A:Accession: A6414

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-403 <BUL>

A:Cross-references: GB:U67535; GB:L77117; NID:g2826348; PIDN:AAB98918.1; PID:g1591587; T

C:Genetics:

A:Map position: REV845792-844581

C:Superfamily: pantothenate metabolism flavoprotein dfp

C:Keywords: flavoprotein

Query Match 2.7%; Score 11; DB 2; Length 403;

Best Local Similarity 100.0%; Pred. No. 0.014;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 SSGKMGVALAE 228

Db 219 SSGKMGVALAE 229

RESULT 5

D81819

N:Alternate names: pantothenate metabolism flavoprotein NMA1916 [imported] - Neisseria meningitidis (strain 22491 serogr

C:Species: Neisseria meningitidis

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: D81819

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A:Reference number: A81775; MUID:20222556

A:Accession: D81819

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <PAR>

A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85137.1; PID:g738

A:Experimental source: serogroup A, strain 22491

C:Genetics:

A:Gene: NMA1916

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 394;

Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 SSGKMGVALA 227

Db 208 SSGKMGVALA 217

RESULT 6

A81058

N:Alternate names: pantothenate metabolism flavoprotein NMB1658 [imported] - Neisseria meningitidis

C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: A81058

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: A81058

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-394 <TET>

A:Cross-references: GB:AF002516; GB:AF002098; NID:g7226905; PIDN:AAF42007.1; PID:g722

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB1658

C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.5%; Score 10; DB 2; Length 394;

Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 SSGKMGVALA 227

Db 208 SSGKMGVALA 217

RESULT 7
D69455
pantothenate metabolism flavoprotein dfp homolog AF1645 - Archaeoglobus fulgidus
N:Alternate names: probable aspartate 1-decarboxylase activase
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: D69455
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: D69455
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-404 <KLE>
A:Cross-references: GB:AE000989; GB:AE000782; NID:q2689312; PIDN:AAB89597.1; PID:g264890
C:Superfamily: pantothenate metabolism flavoprotein dfp

Query Match 2.28; Score 9; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 102 TANTISKIA 110
|||||
DB 103 TANTISKIA 111

RESULT 8
T23443
hypothetical protein K08C9.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T23443
R:Lennard, N.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19741
A:Accession: T23443
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-127 <WIL>
A:Cross-references: EMBL:Z81567; PIDN:CA804589.1; GSPDB:GN00019; CESP:K08C9.6
A:Experimental source: clone K08C9
C:Genetics:
A:Gene: CESP:K08C9.6
A:Map position: 1
A:Introns: 52/2; 69/1
C:Superfamily: Caenorhabditis elegans hypothetical protein K08C9.6

Query Match 2.08; Score 8; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 248 FRIRKIL 255
|||||
DB 85 FRIRKIL 92

RESULT 9
S62355
high mobility group protein 1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 23-Jul-1999
C:Accession: S62355
R:Nightingale, K.; Dimitrov, S.; Reeves, R.; Wolffe, A.P.
EMBO J. 15, 548-561, 1996
A:Title: Evidence for a shared structural role for HMGI and linker histones B4 and H1 in
A:Reference number: S62355; MUID:96174815

A:Accession: S62355
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-210 <NIG>
A:Cross-references: EMBL:U21933; NID:g709958; PIDN:AAC59859.1; PID:g709959
A:Note: the authors did not translate the codon for residue 1
C:Superfamily: nonhistone chromosomal protein HMG-2; HMG box homology
F:6-82/Domain: HMG box homology <HMG1>
F:91-165/Domain: HMG box homology <HMG2>

Query Match 2.08; Score 8; DB 2; Length 210;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 286 SDFRPKIK 293
|||||
DB 106 SDFRPKIK 113

RESULT 10
E71125
Probable aromatic-amino-acid transaminase (EC 2.6.1.57) PH0771 [similarity] - Pyrococ
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: E71125
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137
A:Accession: E71125
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-391 <KAW>
A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29863.1; PID:g3257180
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenB
C:Genetics:
A:Gene: PH0771
C:Superfamily: aspartate transaminase
C:Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate
F:236/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 2.08; Score 8; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 335 TSKEKLIE 342
|||||
DB 371 TSKEKLIE 378

RESULT 11
T36548
Probable transaminase (EC 2.6.1.1-) SCH10.36 [similarity] - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C:Accession: T36548
R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z21609
A:Accession: T36548
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-396 <MUR>
A:Cross-references: EMBL:AL049754; PIDN:CAB42045.1; GSPDB:GN00070; SCOEDB:SCH10.36
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCH10.36
C:Superfamily: aspartate transaminase
C:Keywords: aminotransferase

Query Match 2.0%; Score 8; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 VLVTAGAT 204
Db 99 VLVTAGAT 106
|||||

RESULT 12
TWVTCG
cag phosphoglycerate kinase (EC 2.7.2.3) precursor, chloroplast - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: S05967
R:Longstaff, M.; Raines, C.A.; McMorrow, E.M.; Bradbeer, J.W.; Dyer, T.A.
Nucleic Acids Res. 17, 6569-6580, 1989
A:Title: Wheat phosphoglycerate kinase: evidence for recombination between the genes for
A:Reference number: S05966; MUID:89385983
A:Accession: S05967
A:Molecule type: mRNA
A:Residues: 1-480 <LON>
A:Cross-references: EMBL:X15233; NID:g21832; PIDN:CAA33303.1; PID:g21833
C:Genetics:
A:Map position: 1
C:Superfamily: phosphoglycerate kinase
C:Keywords: ATP; Calvin cycle; chloroplast; phosphotransferase
F:1-72/Domain: transit peptide (chloroplast) #status predicted <TNP>
F:73-480/Product: phosphoglycerate kinase #status predicted <MAT>
F:277,399/Binding site: ATP (Lys, Glu) #status predicted

Query Match 2.0%; Score 8; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 LEGKRVLV 199
Db 86 LEGKRVLV 93
|||||

RESULT 13
B64595
cag pathogenicity island protein cag3 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: B64595
R:Tomb, J.F.; White, O.; Kervlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97594467
A:Accession: B64595
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-481 <TON>
A:Cross-references: GB:AE000566; GB:AE000511; NID:g2313628; PIDN:AAD07589.1; PID:g231363

Query Match 2.0%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 SKEKLIEE 343
Db 281 SKEKLIEE 288
|||||

RESULT 14

D71927
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: D71927
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: D71927
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-481 <ARN>
A:Cross-references: GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD06054.1; PID:g415
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf8

Query Match 2.0%; Score 8; DB 2; Length 481;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 336 SKEKLIEE 343
Db 281 SKEKLIEE 288
|||||

RESULT 15

T36972
probable membrane associated protein - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36972
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21618
A:Accession: T36972
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-794 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CAB52886.1; GSPDB:GN00070; SCOEDB:SCJ11.01c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ11.01c

Query Match 2.0%; Score 8; DB 2; Length 794;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 TAGATREY 207
Db 523 TAGATREY 530
|||||

Search completed: January 31, 2002, 13:20:05
Job time: 106 sec

This Page Blank (uspto)

This Page Blank (uspto)


```

RESULT 2
AAT_PVRHO
ID AAT_PVRHO STANDARD; PRT; 391 AA.
AC OS8489;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1) (TRANSMINASE A) (ASPART).
GN ASPC OR PH0771.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa M., Takamiya M., Onofuku Y.,
RA Funahashi T., Tanaka T., Kudo H. Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shirayama H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: L-ASPARTATE + 2-OXOGLUTARATE -> OXALOACETATE +
CC L-GLUTAMATE
CC -1- COPACITOR; PYRIDOXAL PHOSPHATE (BY SIMILARITY)
CC -1- SIMILARITY: BELONGS TO CLASS-1 OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP000003; BAA29863.1;
CC HSP; Q56232; IBKG.
CC InterPro: IPR001511; AminoTran_1.
CC Pfam: PF00155; aminoTran_1.
CC PROSITE: PS00105; AA_TRANSFER_CLASS_1; 1.
CC TRANSFERASE; AminoTransferase; Pyridoxal phosphate; Complete proteome.
FT BINDING 236 236 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 391 AA; 44565 MW; C35FD73BD08FE4C1 CRC64;

Query Match 2.0%; Score 8; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TSKEKLIE 342
DB 371 TSKEKLIE 378
|||||||

RESULT 3
PGKH_WHEAT
ID PGKH_WHEAT STANDARD; PRT; 480 AA.
AC P12782;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 34, Last annotation update)
DE PHOSPHOGLYCERATE KINASE, CHLOROPLAST PRECURSOR (EC 2.7.2.3).
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CV. MARDLER; TISSUE=Leaf;
RX MEDLINE=89385983; PubMed=2780287;
RA Longstaff M., Raines C.A., McMorrow E.M., Bradbeer J.W., Dyer T.A.;
RT "Wheat phosphoglycerate kinase: evidence for recombination between
RT the genes for the chloroplastic and cytosolic enzymes.";
RL Nucleic Acids Res. 17:6569-6580(1989).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE -> ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: CALVIN CYCLE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X15233; CAA33303.1;
CC PIR: S05967; TWMYGC.
CC HSP; P18912; IHPH.
CC Pfam: PF00162; PGK_1.
CC Pfam: PF00162; PGK_1.
CC PRINTS: PR00477; PHGLYCKINASE.
CC PROSITE: PS00111; GLYCERATE_KINASE; 1.
CC TRANSIT PEPTIDE.
KW TRANSIT PEPTIDE.
FT CHAIN 473 480 CHLOROPLAST (POTENTIAL).
SQ SEQUENCE 480 AA; 49839 MW; 3EBA1F378DAB16CC CRC64;

Query Match 2.0%; Score 8; DB 1; Length 480;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 192 LEGRKRLV 199
DB 86 LEGRKRLV 93
|||||||

RESULT 4
NMT_ASPFU
ID NMT_ASPFU STANDARD; PRT; 492 AA.
AC Q3UVX3;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (EC 2.3.1.97) (PEPTIDE
DE N-MYRISTOYLTRANSFERASE) (MYRISTOYL-COA:PROTEIN N-MYRISTOYLTRANSFERASE)
DE (NMT).
GN NMT1.
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RA Sakata K., Hashido K., Aoki Y., Arisawa M.;
RT "N-mycristoyl-transferase";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADDS MYRISTOYL GROUP TO N-TERMINAL GLYCINE RESIDUE OF
CC CERTAIN CELLULAR PROTEINS.
CC -1- CATALYTIC ACTIVITY: TETRADECANOYL-COA + GLYCYL-PEPTIDE -> COA +
CC N-TETRADECANOYLGLYCYL-PEPTIDE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE NMT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AB035414; BAA87865.1; -
 DR HSSP; P14743; 2NMT.
 DR InterPro; IPR000903; NMT.
 DR Pfam; PF01233; NMT; 1.
 DR PROSITE; PS00975; NMT_1; 1.
 DR PROSITE; PS00976; NMT_2; 1.
 KW Transferase; Acyltransferase.
 SQ SEQUENCE 492 AA; 56255 MW; 921C0C5FF805F000 CRC64;

Query Match 2.0%; Score 8; DB 1; Length 492;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ATKSRKLV 17
 |||||
 DB 183 ATKSRKLV 190

RESULT 5
 PGMU_BROIN STANDARD; PRT; 581 AA.
 AC Q9SNX2;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHOGLUCOMUTASE, CYTOPLASMIC (EC 5.4.2.2) (GLUCOSE PHOSPHOMUTASE)
 DE (PGM).
 GN PGM1.
 OS Bromus inermis (Smooth brome grass).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 CC Bromaeae; Bromus.
 CC NCBI_TaxID=15371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stribitum S., Lee S.P.;
 RT "Expression of a phosphoglucumutase in brome grass suspension culture
 RT cells during abscisic acid induced freezing tolerance."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS ENZYME PARTICIPATES IN BOTH THE BREAKDOWN AND
 CC SYNTHESIS OF GLUCOSE (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ALPHA-D-GLUCOSE 1-PHOSPHATE = ALPHA-D-GLUCOSE
 CC 6-PHOSPHATE.
 CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF197925; AAF04862.1; -
 DR HSSP; P00949; 3PMG.
 DR InterPro; IPR001485; PGM_PMM.
 DR Pfam; PF00408; PGM_PMM; 1.
 DR PRINTS; PR00509; PGM_PMM.
 DR PROSITE; PS00710; PGM_PMM; 1.
 KW Isomerase; Phosphorylation; Magnesium.
 FT ACT_SITE 123 123 FORMS THE PHOSPHOSERINE INTERMEDIATE
 FT (BY SIMILARITY).
 SQ SEQUENCE 581 AA; 62672 MW; 2EDE54521A0F027D CRC64;

Query Match 2.0%; Score 8; DB 1; Length 581;
 Best Local Similarity 100.0%; Pred. No. 6.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 IKEIQPNV 326
 |||||
 DB 477 IKEIQPNV 484

RESULT 6
 EX7S_STRCO STANDARD; PRT; 88 AA.
 ID EX7S_STRCO
 AC Q9FBM4;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE EXODEOXYRIBONUCLEASE VII SMALL SUBUNIT (EC 3.1.11.6)
 DE (EXONUCLEASE VII SMALL SUBUNIT).
 GN XSEB OR SCK7.28C.
 OS Streptomyces coelicolor.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycetaceae; Streptomyces.
 CC NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D., Cerdeno A.M., Parkhill J., Barrell B.G.,
 RA Rajandream M.A.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BIDIRECTIONALLY DEGRADES SINGLE-STRANDED DNA INTO LARGE
 CC ACID-INSOLUBLE OLIGONUCLEOTIDES, WHICH ARE THEN DEGRADED FURTHER
 CC INTO SMALL ACID-SOLUBLE OLIGONUCLEOTIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: EXONUCLEOTIC CLEAVAGE IN EITHER 5'-TO 3'-OR
 CC 3'-TO 5'-DIRECTION TO YIELD 5'-PHOSPHOMONONUCLEOTIDES.
 CC -1- SUBUNIT: HETEROOLIGOMER COMPOSED OF LARGE AND SMALL SUBUNITS (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE XSEB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AL391754; CAC05900.1; -
 KW Hydrolase; Nuclease; Exonuclease.
 SQ SEQUENCE 88 AA; 9376 MW; 3886E0F0CD7EF4FC CRC64;

Query Match 1.7%; Score 7; DB 1; Length 88;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 ALAEAD 231
 |||||
 DB 69 ALAEAD 75

RESULT 7
 Y022_ARCFU STANDARD; PRT; 91 AA.
 ID Y022_ARCFU
 AC O30213;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN AF0022.
 GN AF0022.
 OS Archaeoglobus fulgidus.
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;

```

CC OX Archaeoglobus.
CC NCBI_TaxID=2234;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
CC MEDLINE=98049343; PubMed=9389475;
CC Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
CC Ketchum K.A., Dodson R.J., Gwinn M., Hickie E.K., Peterson J.D.,
CC Richardson D.L., Kerlavage A.R., Graham M.D., Sutton G.G., Gill S.,
CC Fleischmann R.D., Quackenbush J., Lee N.H., Adams M.D., Loftus B.,
CC Kirkness E.F., Dougherty B.A., McKenney K., Badger J.H., Glodek A., Zhou L.,
CC Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
CC Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
CC Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
CC Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
CC Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
CC Venter J.C.;
CC "The complete genome sequence of the hyperthermophilic, sulphate-
CC reducing archaeon Archaeoglobus fulgidus."
CC Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF0022;
CC TIGR: AF0022;
CC Hypothetical protein: Complete proteome.
CC SEQUENCE 91 AA; 10099 MW; E4A2250C4593DAE6 CRC64;
CC -----
CC Query Match 1.7%; Score 7; DB 1; Length 91;
CC Best Local Similarity 100.0%; Pred. No. 14;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 260 VEEMLSA 266
CC | | | | |
CC Db 40 VEEMLSA 46
CC -----
CC RESULT 8
CC ID Y659_METJA STANDARD; PRT; 138 AA.
CC AC Q58073;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DE 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE HYPOTHETICAL PROTEIN MJO659.
CC GN MJO659.
CC OS Methanococcus jannaschii.
CC OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
CC CC Methanococcus.
CC NCBI_TaxID=2190;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
CC MEDLINE=96337999; PubMed=9688087;
CC Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
CC Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
CC Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
CC Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
CC Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
CC Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
CC Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
CC Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
CC "Complete genome sequence of the methanogenic archaeon, Methanococcus
CC jannaschii."
CC Science 273:1058-1073(1996).
CC -1- SIMILARITY: SOME, TO M.JANNASCHII MJEC19.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67513; AAB98658.1;
CC TIGR: MJO659;
CC Hypothetical protein: Complete proteome.
CC SEQUENCE 138 AA; 15470 MW; 96A8AE06757D563C CRC64;
CC -----
CC Query Match 1.7%; Score 7; DB 1; Length 138;
CC Best Local Similarity 100.0%; Pred. No. 19;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 339 KLIIEGK 345
CC | | | | |
CC Db 7 KLIIEGK 13
CC -----
CC RESULT 9
CC ID IF2B_PYRAB STANDARD; PRT; 140 AA.
CC AC Q8UYR6;
CC DT 20-AUG-2001 (Rel. 40, Created)
CC DT 20-AUG-2001 (Rel. 40, Last sequence update)
CC DE 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE PROBABLE TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA).
CC GN EIF2B OR PAR0959.
CC OS Pyrococcus abyssi.
CC OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
CC NCBI_TaxID=29292;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=ORSAY;
CC Heilig R.;
CC "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
CC structure and evolution."
CC Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS
CC BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA
CC (BY SIMILARITY).
CC -1- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA
CC CHAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EIF-2-BETA / EIF-5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ248287; CAB50346.1;
CC InterPro: IPR002735; eIF5_eIF2B.
CC Pfam: PF01873; eIF5_eIF2B; 1.
CC ProDom: PD004078; eIF5_eIF2B; 1.
CC Initiation factor; Protein biosynthesis; Complete proteome.
CC SEQUENCE 140 AA; 16247 MW; 9D40F2C556DB539A CRC64;
CC -----
CC Query Match 1.7%; Score 7; DB 1; Length 140;
CC Best Local Similarity 100.0%; Pred. No. 20;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 191 TLEGRV 197
CC | | | | |
CC Db 76 TLEGRV 82

```

```

RESULT 10
ID IF2B_PVRHO STANDARD; PRT; 140 AA.
AC O58312;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE PROBABLE TRANSLATION INITIATION FACTOR 2 BETA SUBUNIT (EIF-2-BETA).
GN EIF2B OR PH0605
OS Pyrococcus horikoshii.
CC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -1- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS
CC BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA
CC (BY SIMILARITY).
CC -1- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA
CC CHAIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EIF-2-BETA / EIF-5 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: AP000003; BAA29694.1; -
DR InterPro; IPR002735; eIF5_eIF2B.
DR Pfam; PF01873; eIF5_eIF2B; 1.
DR ProDom; PD004078; eIF5_eIF2B; 1.
KW Initiation factor; Protein biosynthesis; Complete proteome.
SQ SEQUENCE 140 AA; 16247 MW; 9D40F2C1428A129A CRC64;

Query Match 1.7%; Score 7; DB 1; Length 140;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 191 TLEGRV 197
Db 76 TLEGRV 82

RESULT 11
ID RL11_RICPR STANDARD; PRT; 145 AA.
AC Q92E24;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 50S RIBOSOMAL PROTEIN L11.
GN RPLK OR RP136.
OS Rickettsia prowazekii.
CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]

```

```

SEQUENCE FROM N.A.
RP STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Scharitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L11P FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: AJ235270; CAAL4504.1; -
DR InterPro; IPR000911; Ribosomal_L11.
DR Pfam; PF00298; Ribosomal_L11; 1.
DR ProDom; PD001367; Ribosomal_L11; 1.
DR PROSITE; PS00359; RIBOSOMAL_L11; 1.
KW Ribosomal protein; RNA-binding; Complete proteome.
SQ SEQUENCE 145 AA; 15312 MW; C327E41CED9F69B0 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 EAATKII 59
Db 125 EAATKII 131

RESULT 12
ID PTP2_NPVOP STANDARD; PRT; 160 AA.
AC O10273;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE 2 (EC 3.1.3.48).
GN PTP-2.
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpNPV).
CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
CC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome."
RL Virology 229:381-399(1997).
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

DR EMBL: U75930; AAC59008.1; -
DR InterPro: IPR000340; DS_phosphatase.
DR InterPro: IPR000387; TYR_phosphatase.
DR Pfam: PF00782; DSPC; 1
DR SMART: SM00195; DSPC; 1
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS00054; TYR_PHOSPHATASE_DUAL; 1.
KW Hydrolase.
FT ACAT-SITE 96 96 BY SIMILARITY.
SQ SEQUENCE 160 AA; 17979 MW; 51DDA804A554301B CRC64;

Query Match 1.7%; Score 7; DB 1; Length 160;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 ECKRVLV 199
Db 88 ECKRVLV 94
|||||||

RESULT 13
ID RM06_DICDI STANDARD; PRT; 170 AA.
AC O21037;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L6.
GN RPL6
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=4468;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RA Iwamoto M., Pi M., Kurthara M., Morio T., Tanaka Y.;
RA MEDLINE=98228272; PubMed=9560439;
RT "A ribosomal protein gene cluster is encoded in the mitochondrial DNA
of Dictyostelium discoideum: UGA termination codons and similarity of
gene order to Acanthamoeba castellanii.";
RL Curr. Genet. 33:304-310(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=AX3;
RC MEDLINE=20279206; PubMed=10821186;
RA Ogawa S., Yoshino R., Angata K., Iwamoto M., Pi M., Kuroe K.,
RA Matsuo K., Morio T., Urushihara H., Yanagisawa K., Tanaka Y.;
RT "The mitochondrial DNA of Dictyostelium discoideum: complete sequence,
gene content and genome organization.";
RL Mol. Genet. 263:514-519(2000).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -!- SIMILARITY: BELONGS TO THE L6P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL: D63523; BAA23574.1;
DR EMBL: AB000109; BAA78082.1;
DR Dictyob; DD07777;
DR InterPro: IPR000702; Ribosomal_L6.
DR InterPro: IPR002358; Ribosomal_L6_1.
DR Pfam: PF00347; Ribosomal_L6; 1.
DR PRINTS: PR00059; RIBOSOMAL_L6.
DR PRODOM: PD002236; Ribosomal_L6; 1.
DR PROSITE: PS00525; RIBOSOMAL_L6_1; FALSE_NEG.

KW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 170 AA; 19083 MW; F2C46222EFBBEB97D CRC64;

Query Match 1.7%; Score 7; DB 1; Length 170;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 SSGKMGV 224
Db 27 SSGKMGV 33
|||||||

RESULT 14
ID DSBX_XYLFA STANDARD; PRT; 173 AA.
AC Q9PGG2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE).
GN DSBX OR XF0340.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A., Camargo L.E.A., Carraro D.M., Carier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Reto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfy H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.L.,
RA Fraga J.S., Franca S.C., Franco M.C., Fromme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.C.R., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000)
CC -!- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
CC PERIPLASMIC PROTEINS. ACTS BY OXIDIZING THE DSBX PROTEIN (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DSBX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; AE003886; AAF83150.1; -
DR InterPro; IPR003752; Dsbb.
DR Pfam; PF02600; Dsbb; 1.
KW Oxidoreductase; Redox-active center; Electron transport; Chaperone;
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 12 32. POTENTIAL.
FT TRANSMEM 40 60 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT DISULFID 38 41 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 104 131 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 173 AA; 19231 MW; 626DF05E3FB3A49C CRC64;

Query Match 1.7%; Score 7; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 328 LVGFKAE 334
| | | | |
Db 160 LVGFKAE 166

RESULT 15
CLP1_CHLMU STANDARD; PRT; 191 AA.
AC Q9PJW1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT 1 (EC 3.4.21.92)
DE (ENDOPEPTIDASE CLP 1).
GN CLP1 OR TC0715.
OS Chlamydia muridarum
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=833560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: CLP CLEAVES PEPTIDES IN VARIOUS PROTEINS IN A PROCESS
CC THAT REQUIRES ATP HYDROLYSIS. CLP MAY BE RESPONSIBLE FOR A FAIRLY
CC GENERAL AND CENTRAL HOUSEKEEPING FUNCTION RATHER THAN FOR THE
CC DEGRADATION OF SPECIFIC SUBSTRATES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS TO SMALL PEPTIDES IN
CC THE PRESENCE OF ATP AND MAGNESIUM. ALPHA-CASEIN IS THE USUAL TEST
CC SUBSTRATE. IN THE ABSENCE OF ATP, ONLY OLIGOPEPTIDES SHORTER THAN
CC FIVE RESIDUES ARE CLEAVED (SUCH AS SUCCINYL-LEU-TYR-|-NHMEC; AND
CC LEU-TYR-LEU-|-TYR-TRP, IN WHICH THE CLEAVAGE OF THE -TYR-|-LEU-
CC AND -TYR-|-TRP-BOND ALSO OCCURS).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S14; ALSO KNOWN AS CLPP
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002340; AAF39527.1; ALT_INIT.
DR TIGR; TC0715; -
DR InterPro; IPR001907; CLP_protease.
DR Pfam; PF00574; CLP_protease; 1.

DR PRINTS; PR00127; CLPPROTEASEP.
DR PROSITE; PS00381; CLP_PROTEASE_SER; FALSE_NEG.
DR PROSITE; PS00382; CLP_PROTEASE_HIS; 1.
KW Hydrolase; Serine protease; Complete proteome.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 91 91 PROBABLE.
FT ACT_SITE 116 116 PROBABLE.
SQ SEQUENCE 191 AA; 20946 MW; 30196D0814C2339E CRC64;

Query Match 1.7%; Score 7; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 117 PVTTVVT 123
| | | | |
Db 80 PVTTVVT 86

Search completed: January 31, 2002, 13:39:12
Job time: 72 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:12:34 ; Search time 35.08 Seconds
(without alignments)
1680.381 Million cell updates/sec

Title: US-08-957-709-19
Perfect score: 1998
Sequence: 1 MLHHVKLIYATKSRKLVGKK.....KMKRELAERIWDEIEKXLS 403

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_17:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1667.5	83.5	401	1 O59114	O59114 pyrococcus
2	1649.5	82.6	401	1 O9V0S3	O9V0S3 pyrococcus
3	735	36.8	404	1 O28628	O28628 archaeoglob
4	731	36.6	382	1 O9HH70	O9HH70 methanobact
5	731	36.6	386	1 O27284	O27284 methanobact
6	626	31.3	392	1 O9HRS1	O9HRS1 halobacteri
7	535.5	26.8	388	2 O66997	O66997 aquifex ao
8	535	26.8	437	1 O9YA10	O9YA10 aeropyrum p
9	494.5	24.7	404	2 O9K9Y4	O9K9Y4 bacillus ha
10	490.5	24.5	399	2 O9KVD1	O9KVD1 vibrio chol
11	485	24.3	399	2 O9RQH7	O9RQH7 listeria mo
12	483.5	24.2	402	2 O9HTN4	O9HTN4 pseudomonas
13	482.5	24.1	394	2 O9X213	O9X213 thermotoga
14	473.5	23.7	406	2 O35033	O35033 bacillus su
15	473	23.7	394	2 O9JVB7	O9JVB7 neisseria m
16	464	23.2	376	2 O9KX52	O9KX52 streptomyce
17	464	23.2	394	2 O9JY98	O9JY98 neisseria m
18	454.5	22.7	417	2 O9X3X4	O9X3X4 zymomonas m
19	453.5	22.7	399	2 O99UQ7	O99UQ7 staphylococ

20	447	22.4	422	2 O9RWT6	O9RWT6 deinococcus
21	425.5	21.3	394	1 O9HJS5	O9HJS5 thermoplas
22	418.5	20.9	419	2 O9CCQ5	O9CCQ5 mycobacteri
23	405.5	20.3	400	2 O9CLR0	O9CLR0 pasteurella
24	404.5	20.2	409	2 O9PGZ7	O9PGZ7 xylolla fas
25	397	19.9	412	2 O9AJ54	O9AJ54 caulobacter
26	351	17.6	425	2 O9ZL03	O9ZL03 helicobacte
27	350	17.5	425	2 O25512	O25512 helicobacte
28	348.5	17.4	444	2 O52596	O52596 bradyrhizob
29	301	15.1	384	2 O9PPA1	O9PPA1 campylobact
30	237	11.9	228	2 O99Z11	O99Z11 streptococc
31	232	11.6	181	2 O99Z10	O99Z10 streptococc
32	224.5	11.2	231	2 O9CHZ5	O9CHZ5 lactococcus
33	223	11.2	178	2 O9CHZ6	O9CHZ6 lactococcus
34	180	9.0	183	12 O9J5A8	O9J5A8 fowlpox vir
35	176.5	8.8	625	3 O9UT17	O9UT17 schizosacch
36	170	8.5	289	12 O9J5E0	O9J5E0 fowlpox vir
37	169	8.5	277	5 O9BL36	O9BL36 caenorhabdi
38	168.5	8.4	174	2 O9X4Q1	O9X4Q1 streptomyce
39	161.5	8.1	329	11 O9D376	O9D376 mus musculu
40	160	8.0	284	4 O9HAB8	O9HAB8 homo sapien
41	158	7.9	188	2 O9FDM9	O9FDM9 streptococc
42	157	7.9	188	2 O9RPL6	O9RPL6 streptococc
43	153	7.7	188	2 O9F0G2	O9F0G2 streptococc
44	150	7.5	127	4 O9HC17	O9HC17 homo sapien
45	149.5	7.5	270	10 O9LZM3	O9LZM3 arabidopsis

ALIGNMENTS

RESULT 1					
O59114					
ID O59114	PRELIMINARY;	PRT;	401 AA.		
AC O59114					
DT 01-AUG-1998	(TRENBLREL. 07, Created)				
DT 01-AUG-1998	(TRENBLREL. 07, Last sequence update)				
DT 01-JUN-2001	(TRENBLREL. 17, Last annotation update)				
DE 401AA	LONG HYPOTHETICAL PROTEIN DFP.				
GN PH1444					
OS Pyrococcus horikoshii.					
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.					
OX NCBI_TaxID=53953;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=OT3;					
RX MEDLINE=98344137; PubMed=9679194;					
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,					
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,					
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,					
RA Funahashi T., Tanaka T., Kudo H. Y., Yamazaki J., Kishida N., Oguchi A.,					
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,					
RA Masuchi Y., Shizuya H., Kikuchi H.;					
RT *Complete sequence and gene organization of the genome of a hyper-					
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";					
RL DNA Res. 5:55-76(1998).					
DR EMBL: AP000006; BAA30551.1; -					
DR InterPro: IPR003382; Flavoprotein.					
DR Pfam: PF02441; Flavoprotein; 1.					
KW Complete proteome.					
SQ SEQUENCE 401 AA; 44624 MW; 26CBC5F523AE02C8 CRC64;					
Query Match	83.5%;	Score	1667.5;	DB 1;	Length 401;
Best Local Similarity	83.4%;	Pred. No.	3e-82;		
Matches	337;	Conservative	33;	Mismatches	29;
				Indels	5;
				Gaps	4;
OY 1	MLHHVKLIYATKSRKLVGKKVXXXXXPGSIAALD-VKACEGLIRHGAHVAVMSEATKII	59			
Db 1	MLHHVKLIYATKSRKLVGKKVXXXXXPGSIAALD-VKACEGLIRHGAHVAVMSEATKII	60			
OY 60	HPYAWNLPNGNVPITETGTFIEHVELAGEHENKADLLVCPATANTISKIACGIDDPVPT	119			

Db 61 HPYAMEFATGNPVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 120
Qy 120 TVVTTAFPHIPIMIAPIAMHETMYRHPVIRENIERLKKLVGEFIPRIEGRKAKVASIDEI 179
Db 121 TVVTTAFPHIPIMIAPIAMHESMYKHPVIRENIERLKKLVGEFIPRIEGRKAKVASIDEI 180
Qy 180 VYRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSCKMGVALAEAEADFRGA-VTL 238
Db 181 VYRVIRKLHPKTLGKRVLTAGATREYIDPIRFTNASSCKMGVALAEAEAEFRGAETL 240
Qy 239 IRTGSKVAFIRIKLKVETVEEMLSAIENELSKYDVVIMAAVSDFRPKIKAEKKI 298
Db 241 IRTGSKVSNFVENQI--QVETVEEMLSAIENELSKYDVVIMAAVSDFRPKIKAEKKI 298
Qy 299 KSGSITIELVPXNPXIIRIKETQPNVFLVGFKAETSKKELIEEGKQIERAKDLVWG 358
Db 299 KSNKSIITIELVP-NPKIIRIKETQSDVFLVGFKAETSMKELISEAKKQIESAGSDLVIG 357
Qy 359 NTLEAFGSEENOVVLIGRDTFKELPKMKKRELAERIWEIDEIEKXL 402
Db 358 NTLEAFGSEENOVVLIGRDTFKELPKMKKRELAERIWEIDEIEKRL 401
RESULT 2
Q9V0S3 PRELIMINARY; PRT; 401 AA.
AC Q9V0S3
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN (DFP).
GN PAB1897.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution." to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248285; C849630.1;
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Complete proteome.
SQ SEQUENCE 401 AA; 44290 MW; 96CCD5A190C1FA5D CRC64;

Query Match 82.6%; Score 1649.5; DB 1; Length 401;
Best Local Similarity 82.9%; Pred. No. 2.8e-81;
Matches 335; Conservative 31; Mismatches 33; Indels 5; Gaps 4;

Qy 1 MLHVKLIYATKSRKLVGKKTIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKII 59
Db 1 MSHIKLIYATKSRKLVGKKTIVLAPGSIAAVECVKLARELIRHGAEVHVMSPSATKII 60
Qy 60 HPYAWNLTGNPVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 119
Db 61 HPYAMEFATGNPVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 120
Qy 120 TVVTTAFPHIPIMIAPIAMHETMYRHPVIRENIERLKKLVGEFIPRIEGRKAKVASIDEI 179
Db 121 TVVTTAFPHIPIMIAPIAMHESMYKHPVIRENIERLKKLVGEFIPRIEGRKAKVASIDEI 180
Qy 180 VYRVIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSCKMGVALAEAEADFRGA-VTL 238
Db 181 VYRVIRKLHPKTLGKRVLTAGATREYIDPIRFTNASSCKMGVALAEAEAEFRGAETL 240
Qy 239 IRTGSKVAFIRIKLKVETVEEMLSAIENELSKYDVVIMAAVSDFRPKIKAEKKI 298
Db 241 IRTGSKVSNFVENQI--EVEETVEEMLSAIENELSKYDVVIMAAVSDFRPKIKAEKKI 298

Qy 299 KSGSITIELVPXNPXIIRIKETQPNVFLVGFKAETSKKELIEEGKQIERAKDLVWG 358
Db 299 KSDKSIITIELVP-NPKIIRIKETQSDVFLVGFKAETSMKELISEAKKQIESAGSDLVIG 357
Qy 359 NTLEAFGSEENOVVLIGRDTFKELPKMKKRELAERIWEIDEIEKXL 402
Db 358 NTLEAFGSEESKVVIVGKDFVKELPKMKKRELAERIWEIDEIEKI 401
RESULT 3
O28628 PRELIMINARY; PRT; 404 AA.
AC O28628
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PANTOTHENATE METABOLISM FLAVOPROTEIN (DFP).
GN AFI645.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OX Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RA MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 350:364-370(1997).
DR EMBL; AE000969; AAB89597.1;
DR TIGR; AFI645;
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 404 AA; 44885 MW; E5A2B899C3B6A66F CRC64;

Query Match 36.8%; Score 735; DB 1; Length 404;
Best Local Similarity 45.0%; Pred. No. 3.7e-32;
Matches 182; Conservative 74; Mismatches 122; Indels 26; Gaps 13;

Qy 4 HVKLIYATKSRKLVGKKTIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIIHPY 62
Db 5 HLERIRGRSRKLVGKKTIVLGVGTGSIAAVETVKLARELVRRGADVTAVMSRAARKIIHPY 64
Qy 63 AWWLPTGNPVITEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 122
Db 65 ALEFATCKRVVTEITGTFIEHVELAGEHENKADLILVCPATANTISKIACGIDDPVT 123
Qy 123 TPAF-PHIPIMIAPIAMHETMYRHPVIRENIERLKKLVGEFIPRIEGRKAKVASIDEI 181
Db 124 TPAF-PHIPIMIAPIAMHETMYRHPVIRENIERLKKLVGEFIPRIEGRKAKVASIDEI 183
Qy 182 RVYIKLHKKTLEGRKVLVTAGATREYIDPIRFTNASSCKMGVALAEAEADFRGA-VTL 240
Db 184 HVERELTPKMKKRVVVTGPTVEQIDPIRFTNASSCKMGVALAEAEADFRGA-VTL 243
Qy 241 TKSVAKAFIRIKLKVETVEEMLSAIENELSKYDVVIMAAVSDFRPKIKAEKKI 300
Db 244 SKPSGMSLPNVK-EIRVNSVEDMMKAVLYEI-GRGCDLFFVSSAAADFIYDAEAK-KIKT 300
Qy 301 GRSTTIELVPXNPXIIRIKETQPNVFLVGFKAET--SKEKLEEGKQIERAKDLVWG 358

Db 301 APELVKIL-KESPKIIKVRKISG-HIIGFKAETGMSDDLLKVASMKADDNLNVVA 358
Qy 359 NTL--EAFGSENVVLIGRDTFRELPMK-----KRELAERI 394
Db 359 NDVLERGMDTDRVLIL-----TPKQEWVGLKOHVAERI 395

RESULT 4

Q9HH70 ID Q9HH70 PRELIMINARY; PRT; 382 AA.
AC Q9HH70
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MTW1216.
GN Methanobacterium wolfeii.
OS Methanobacterium wolfeii.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145261;
RN [1]
RP SEQUENCE FROM N.A.
RA Luo Y., Pfister P., Leisinger T., Wasserfallen A.;
RT "The genome of archaeal prophage psiM100 encodes the lytic enzyme
RL responsible for autolysis of Methanothermobacter wolfeii.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF301375; AAG39941.1;
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
SQ SEQUENCE 382 AA; 41207 MW; 8E792042DFDFA3AB CRC64;

Query Match 36.6%; Score 731; DB 1; Length 382;

Best Local Similarity 45.5%; Pred. No. 5.6e-32;

Matches 178; Conservative 65; Mismatches 126; Indels 22; Gaps 13;

Qy 20 KIVXXPGSIAALD-VKACEGLIRHGAHVMSAAATKIIHPYAWNLPTGNPVITEITG 78
Db 2 RVLCVTGSAVAEAVKLARELRKGAECVCFMSEDAKRIIHPYAMEFATGSPVLELTG 61
Qy 79 FIEHVELAGEHENKADILVCPATANTISKIACIDIDTPVTVTVTAFPH-IPMIAPAM 137
Db 62 EIEHVKY-----SDADILVAPATANTIGKAYLADNPISLLLTAAAGRTAILMVP 116
Qy 138 HETMYRHPVIRENTERLKLGVFIEGRIEGRKAVASIDIVYRVTKLHKKTLEGRV 197
Db 117 HEAMY--AAENRKLKEGVAFLEPRMDEGKAKFPDIDITILELRQASEGKMRGLRV 174
Qy 198 LVTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VTLIRTKGSV-AFRIRKIKL 255
Db 175 LVSLGGTLEPIDVRVTNRSSGRMLAVAREAVIQGADVTLVAGTVSDIPSLRTV-- 232
Qy 256 KVEIVTEMLSAIENELSKKYDVVIMAAVSDFRPKIKAEKIKSGRSITIELVXPNKI 315
Db 233 RAETAHEMAEVAELI--GEHDVFVSAVSDFRP-VYSEKISSDSEITLRLKP-NPKI 288
Qy 316 IDRKEIQPNVFLVGFKAEE--TSKEKLEIEGKROIERAKADLVVGN--TLEAFGSENOV 371
Db 289 IRMARETNPEAFIVGFAEHGVSEELIAARKQIEDSVADMVANDVSVGFGSENNRA 348
Qy 372 VLIGRDTFKELPKMKKRELAERIWDEIEKXL 402
Db 349 IIVSEGVY-ELPTMKKEELAGLIIGIEIMKRL 378

RESULT 5

O27284 ID O27284 PRELIMINARY; PRT; 386 AA.
AC O27284
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE PANTOTHENATE METABOLISM FLAVOPROTEIN.
GN MTW1216.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE-98037514; PubMed-9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., DuBois J.,
RA Alredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000889; AAB85705.1;
DR InterPro: IPR003382; Flavoprotein.
DR Pfam: PF02441; Flavoprotein; 1.
RW Complete proteome.
SQ SEQUENCE 386 AA; 42010 MW; 56386EAB76244A75 CRC64;

Query Match 36.6%; Score 731; DB 1; Length 386;

Best Local Similarity 44.9%; Pred. No. 5.7e-32;

Matches 176; Conservative 62; Mismatches 122; Indels 32; Gaps 13;

Qy 20 KIVXXPGSIAALD-VKACEGLIRHGAHVMSAAATKIIHPYAWNLPTGNPVITEITG 78
Db 2 EILCVTGSVAIEAVKLARELRKGAECVCFMSEDAKRIIHPYAMEFATGSKPVLELTG 61
Qy 79 FIEHVELAGEHENKADILVCPATANTISKIACIDIDTPVTVTVTAF-PIHIPMIAPAM 137
Db 62 EIEHVKYAG-----ADILVAPATANTIGKAYLADNPISLLLTASGMDTPIVMVPSM 116
Qy 138 HETMYRHPVIRENTERLKLGVFIEGRIEGRKAVASIDIVYRVTKLHKKTLEGRV 197
Db 117 HEAMY--AAENRKLKEGVFIEPRMDEGKAKFPDIDITILELRQASEGKMRGLRV 174
Qy 198 LVTAGATREYIDPIRFTNASSGKMGVALAEADFRGA-VTLIRTKGSV-----KAFRI 250
Db 175 LVSLGGTLEPIDVRVTNRSSGRMLAARRAVIEGADVTLVAGTVSVEIPPLRSFR- 233
Qy 251 RKIKLVETVEMLSAIENELSKKYDVVIMAAVSDFRPKIKAEKIKSGRSITIELV 310
Db 234 -----AETAEMAERVELVAD--HDVFISSAAVADFKP-VYTERKISSSEFSEVLRP 284
Qy 311 XNPKIIDRIKEIQPNVFLVGFKAEE--TSKEKLEIEGKROIERAKADLVVGN--TLEAFGS 366
Db 285 -NPKVTIGAREINPEAFIVGFAEYVDNDEALVESARKQIRESGVDMVANDVSVGFGS 343
Qy 367 EENQVLIIGRDTFKELPKMKKRELAERIWDEI 398
Db 344 DRNRALIVS-DMVTLPMEKEELASIIIDEV 374

RESULT 6

Q9HRS1 ID Q9HRS1 PRELIMINARY; PRT; 392 AA.
AC Q9HRS1
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PANTOTHENATE METABOLISM FLAVOPROTEIN.
GN DFP OR VNG0572G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;

KW Complete proteome.
SQ SEQUENCE 437 AA; 46842 MW; 91463879F31F1618 CRC64;

Query Match 26.88; Score 535; DB 1; Length 437;
Best Local Similarity 36.88; Pred. No. 2.3e-21;
Matches 150; Conservative 66; Mismatches 152; Indels 40; Gaps 14;

QY 18 GKIVXXPGSIA---ALDVKACEGLIRHGAEVHVMSEAATKIHPYAWNLPTGNPVIT 74
DB 47 KTIIVGLVTSVALYRSIDL--ARWLLRGARVITVMTPEAKLVSPMFWASGGPVIT 104

QY 75 BITGFIEHVELAGHENKADLILVCPATANTISKIACGIDDPVTVVTTAPPH-IPIMI 133
DB 105 GFTGVEHISIA---RAASAMVAVPATLSTLAKIAHGVDNPNVALAAVSTMGYKPVIA 160

QY 134 APAMHETMYRPIVRENIERLKLGVFEIGPRIEGRKAKVASIDEIVRVYIKLHK--KT 191
DB 161 YPMHGNMYESPQAREVVDRLRSQGLVVDPKIEGGVAKYPTDHAVGRITAAQARKGLRD 220

QY 192 LEGRVLVTAGATREYIDPIRFITNASSGKMGVALAEADPRGA-VTLIRTKGSKVAFRI 250
DB 221 LEGIRALVTGLSTREWDIDRFISNPSSGVGMLEALELYARGAEVDVAGYTSVEIPHL 280

QY 251 RKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAEKIKSGRSITIELVP 310
DB 281 FN-TVKTETEDMAAAVEELTSKREYDAVVAAPVDFRPAAGAFEGKIRSGQLVLELEP 339

QY 311 XNPKLIIDRIKIQPNVFLVGFKAET--TSKEKLIIEGKRQIERAKADLVVGNLTLEAFG--- 365
DB 340 -TPKVLGEGIAR-RPKV-LVAPAAEYVDNLSRDPALPEKMEKYDADLVVANRVGEGVGF 396

QY 366 -----SEENQVVLIDGFTKELPKMKKRELAERIDEIEKXLS 403
DB 397 ASPLDLVLMDKSEAVLKG-SFKHEI-----VAAVIADEIAKLKS 436

RESULT 9
Q9K9Y4 ID Q9K9Y4 PRELIMINARY; PRT; 404 AA.
AC Q9K9Y4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE FLAVOPROTEIN.
GN DFP OR BH2510.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001515; BAB06229.1; -
DR HSP; Q9SWES; 1E20.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Complete proteome.
SQ SEQUENCE 404 AA; 44044 MW; 47FF382F2AEB9E7C CRC64;

Query Match 24.78; Score 494.5; DB 2; Length 404;
Best Local Similarity 33.58; Pred. No. 3.1e-19;
Matches 138; Conservative 84; Mismatches 149; Indels 41; Gaps 15;

QY 16 LVGKKIVXXPGSIAALDVKA-CBGLIRHGAEVHVMSEAATKIHPYAWNLPTGNPVIT 74

DB 2 LQGRVVLGVSGGIAAFKSAFAASKLVQAGAEVAVVMTGAKKFVPLTFOALTRHPVD 61

QY 75 EL-----TGFIHVELAGHENKADLILVCPATANTISKIACGIDDPVTVVTTAPPHI 129

DB 62 DTFSEPDSEIAHQLA-----DWADVIIATPATANLIGKLANGVADMLSTMLLAT--KA 115

QY 130 PIMTAPAMHETMYRPIVRENIERLKLGVFEIGPRIEE-----GRAKVASIDEIVRV 183

DB 116 PIYLAPAMVNMVYEPVORNMQQLAKDGYRLLEPGAGYLACGWIGRMPEDLLKTI 175

QY 184 IKLH-----KTKLGKRVLTAGATREYIDPIRFITNASSGKMGVALAEAE-DEFGAVTL 238

DB 176 --EVHFTPPSSSLAGKKIVITAGPTQETIDPIREFTNRSSGKMGYALTKAARDFGNVTL 233

QY 239 IRTKGSV-KAFIRIKIKLVETVEEMLSAIENELRSKKYDVVIMAAVSDFRPKIKAE 297

DB 234 ISGTPSLEKPDGVSVMVK--SAQDMYEAFLAEFSDA--DVVIKTAADVADRVRRVHEQK 289

QY 298 IKSGRSITIELVXPNPKLIIDRIKIQPNVFLVGFKAETSKELIEGKRQIERAKADLVV 357

DB 290 VKKDGWDVIELERTVDILKTLGKKESQFLVGFPAES--QEVETYAKKLKERNADMIV 347

QY 358 GNTL-----EAFGSEENV-VLIGRDFTKELPKMKKRELAER-----IWDEIEK 400

DB 348 ANNVTEGAGFTDITNRVTVYFKQGDVKKPLPLMTKDEVAHRLIMMISQLEK 399

RESULT 10
Q9KVD1 ID Q9KVD1 PRELIMINARY; PRT; 399 AA.
AC Q9KVD1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN.
GN VC0215.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004111; AAF93391.1; -
DR HSP; Q9SWES; 1E20.
DR TIGR; VC0215; -
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Complete proteome.
SQ SEQUENCE 399 AA; 42645 MW; 65192A744F857028 CRC64;

Query Match 24.58; Score 490.5; DB 2; Length 399;
Best Local Similarity 35.18; Pred. No. 5e-19;
Matches 145; Conservative 69; Mismatches 160; Indels 39; Gaps 15;

QY 14 RKLVGKKIVXXPGSIAALD-VKACEGLIRHGAEVHVMSEAATKIHPYAWNLPTGNPV 72

DB 2 QSLAGKKILIGISGGIAAYKCAELTRRLVGRGATVQVVMTHAAKEFTPLTMQAVSRPV 61

QY 73 ITEI-----TGFIHVELAGHENKADLILVCPATANTISKIACGIDDPVTVVTTV 125

```
Db 62 SDSLLDPAEASMGHIELA----KWADLVLLAPATADLIARMAAGMGNDLLTTLATSA 117
Qy 126 PPHIPIMAPAMHETMYRHPVIRENIERKLKLGVEFFIGRIE-----GRAKVASIDEI 179
Db 118 ----PVAIPAPAMNOOMYRNIAATOENLOTLIRRGYLTWGPAGEACGVDGPGRMLEPMEL 173
Qy 180 VYRVIKLHKHTLEGRKVLVTAGATREVIDPTREFTTNASSGKMGVALAEADFRGA-VTL 238
Db 174 VAHCENFFAPKILVGKRVLTITAGPTREALDPVRYITNHSSGKMGFALAKAAQAQLGADVTL 233
Qy 239 IRTGSKV-AFRIRKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDRFPKIKAGK 297
Db 234 V--SGPVHLPTPGVNRIDVQSGLEHSAVMKEATS--HQIFIACAADVAPQVTAEQK 289
Qy 298 IKSGR---SITIELVPXNPKIIDRIKEIQPN-VFLVGFKAEKSKELIEGKROIERAKA 353
Db 290 IKKSNDNLTLEIEMW-KNPDIVASVAALTENRPFTVGAETQDVETVARSK--LVRKNL 346
Qy 354 DLVVCNTLE---EAFSGSENOVVLIGRDTKELPKMKKRELAERIWDIEKXL 402
Db 347 DMICANDYSIAGOGFNSNDNALTLEWKGQHSPLTSLTKDALASAVMHLIHEQM 399

RESULT 11
Q9RQH7 PRELIMINARY; PRT; 399 AA.
AC Q9RQH7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG.
GN DFP.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP STRAIN=EGD;
RC "Isolation and characterization of Listeria monocytogenes mutants altered for adherence to eucaryotic cells."
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF104226; AAF04763.1; -
DR HSSP; Q9SWE5; 1E20.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
SQ SEQUENCE 399 AA; 43395 MW; 566E98FF6F07F7EB CRC64;

Query Match 24.3%; Score 485; DB 2; Length 399;
Best Local Similarity 34.8%; Pred. No. 9.8e-19;
Matches 144; Conservative 69; Mismatches 133; Indels 48; Gaps 16;
Qy 18 GKIVYXXPGSIALDVKACEG-LIRHGAEVHVMSEAAATKIIHPYANLPTGNPVITEI 76
Db 3 GRNILLAVSGGLVYKVALTSKLTQAGANVKVMHTAHAEFVPLSPQVLSKNDVYDT 62
Qy 77 -----TGIEHVELAGHENKADLIVCPATANTISKIACGIDPTPTVTVTTAPPHIPI 131
Db 63 FDEKSSVVAHIDLA----DWADLVIVAPATANYIGNKANGIADDMVTITLAT--EAPV 116
Qy 132 MIAPAMHETMYRHPVIRENIERKLKLGVEFFIGRIE-----GRAKVASIDEIVRV 183
Db 117 WVAPAMNYHMIOHPAVIRINRLNADGVRFTEP--EEGLACGVYVGRGRLEEPEKIVLR 174
Qy 184 IK--KLHKKTLEGRKVLVTAGATREYIDPTREFTTNASSGKMGVALAEADFRGA-VTLR 240
Db 175 AEFFQEDKNLQGNVLTAGATREKLDLIRYFTNHSTGKMGFSAESAAARHGANTVLT 234
Qy 241 TKGS-----VKAFRIRKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDRFPKIK 294
Db 235 TSKALPVPHGVEA-----IYVESAEHQAQV-NE-RKVSODIFVMTAAVADYTPAQVS 285
```

```
Qy 295 EGKI-KSGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEKSKELIEGKROIERAKA 353
Db 286 DQIKKQPGDEFTIAMKRTKIDILLEGQHKTSQVVGFAAET--ENVEANARKKLTSKNA 343
Qy 354 DLVVCNTLE---AFSGSENOVVLIGRDTKRE-LPKMKKRELAERIWDIEKXL 402
Db 344 DMIVANNISEAGAGSGDTNIVTFYRKDGSSEALPILDKKEVAEHIKEAANFL 397

RESULT 12
Q9HTN4 PRELIMINARY; PRT; 402 AA.
AC Q9HTN4
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN.
GN DFP OR PA5320.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004944; AAG08705.1; -
DR HSSP; Q9SWE5; 1E20.
DR InterPro; IPR003382; Flavoprotein.
DR Pfam; PF02441; Flavoprotein; 1.
KW Complete proteome.
SQ SEQUENCE 402 AA; 43133 MW; 5F1CE1CBEC2B67D3 CRC64;
```

```
Query Match 24.2%; Score 483.5; DB 2; Length 402;
Best Local Similarity 33.4%; Pred. No. 1.2e-18;
Matches 140; Conservative 81; Mismatches 147; Indels 51; Gaps 17;
```

```
Qy 14 RKLVGKTIYXXPGSIALDVKACEGLIR---HGAEVHVMSEAAATKIIHPYANLPTG 69
Db 2 QRLYRKRLVLGVGGGIAA--YKSAB-LVRLRDQGAEVVYVMTGGREFIPLTLQALSG 58
Qy 70 NPVITEI-----TGIEHVELAGHENKADLIVCPATANTISKIACGIDPTPTVTVTT 124
Db 59 HPVHTDLLDPAEASMGHIELA---RWADLVLIAPATADLMARLVQGVANDLLTTLVLA 114
Qy 125 AFPHIPIMAPAMHETMYRHPVIRENIERKLKLGVEFFIGPRI-----EGRKVASIDE 178
Db 115 T--DAQIALAPAMNOAMWRDQTATQANELLRQGFHLFGPAAGSQACGDVGLGRMLEAE 172
Qy 179 IYRVIRKHKHTLEGRKVLVTAGATREYIDPTREFTTNASSGKMGVALAEADFRGA-VT 237
Db 173 LAQRAADCQFQALTGTVHVLITAGPTQENIDPVRYITNHSSGKMGFALAEAAVEAGARV 232
Qy 238 LIRTKGSV---KAFRIRKIKLVETVEMLSAIENELRSKKYDVVIMAAVSDRFPKIK 294
Db 233 LV--TGPVHLPTDPRQVR--DVVSARDMLAAE---AAMPDLLIASAAVADYRPEVA 285
Qy 295 BGKIK---SGRSITIELVPXNPKIIDRIKEIQPNVFLVGFKAEKSKELIEGKROIER 350
Db 286 AHKLKQPTSGEGLLQLV-RNPDLATLAQREDRPFVSGFAAET--ENLDYAAARKLD 342
Qy 351 AKADLVVGNLTLE---AFSGSENOVVLIGRD-----FTKELPKMKKRELAERIWDIEK 400
```


This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:08:39 ; Search time 21.39 seconds
(without alignments)
555.551 Million cell updates/sec

Title: US-08-957-709-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKEILIEPSEEE.....PYRGNVQSGTRLAFSKRKKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	768	95.3	156	2 E71216	dCTP deaminase (EC
2	703	87.2	154	2 G75030	dCTP deaminase (EC
3	243	30.1	180	2 A70439	probable dCTP deam
4	242.5	30.1	177	2 H83695	deoxycytidine triph
5	242	30.0	173	2 S26382	probable dCTP deam
6	232.5	28.8	181	2 D72724	probable dCTP deam
7	216	26.8	191	2 T36613	probable dCTP deam
8	199	24.7	200	2 A69114	dCTP deaminase (EC
9	198	24.6	190	2 B70526	dCTP deaminase (EC
10	189.5	23.5	186	2 A81272	probable dCTP deam
11	184.5	22.9	195	2 G84184	deoxycytidine triph
12	181	22.5	190	1 D64566	dCTP deaminase (EC
13	180	22.3	188	2 E71860	dCTP deaminase (EC
14	174	21.6	206	2 C84942	dCTP deaminase (EC
15	172	21.3	188	2 A83210	probable deoxycyti
16	165	20.5	193	2 C85833	2'-deoxycytidine 5
17	164.5	20.4	191	2 G82765	deoxycytidine triph
18	163	20.2	193	1 A42940	dCTP deaminase (EC
19	159	19.7	163	2 H72759	probable dCTP deam
20	155	19.2	188	2 D81149	deoxycytidine triph
21	154	19.1	204	2 F64353	dCTP deaminase (EC
22	153	19.0	150	2 D69081	deoxyuridine 5-tri
23	150.5	18.7	190	2 D81717	deoxycytidine triph
24	149.5	18.5	190	2 B71565	probable dCTP deam
25	149.5	18.5	195	1 A64050	dCTP deaminase (EC
26	145	18.0	188	2 E71715	probable dCTP deam
27	138	17.1	190	2 F86539	dCTP deaminase [im
28	138	17.1	190	2 F72084	dCTP deaminase (EC
29	137	17.0	172	2 T44356	probable dCTP deam

30 135.5 16.8 193 1 S75588 dCTP deaminase (EC
31 126 15.6 168 2 C69388 probable dCTP deam
32 125.5 15.6 1145 1 GNLJEV pol polyprotein -
33 125.5 15.6 1146 1 GNLJEV pol polyprotein (c
34 125.5 15.6 1146 1 GNLJ22 pol polyprotein -
35 118 14.6 1109 1 B45345 pol polyprotein -
36 117.5 14.6 161 2 E64437 probable dCTP deam
37 115.5 14.3 188 3 JC7565 nucleoside-triphos
38 111.5 13.8 1086 1 B46335 pol polyprotein -
39 111 13.8 141 2 A46256 dCTP pyrophosphata
40 111 13.8 164 2 G02777 dCTP pyrophosphata
41 111 13.8 1101 1 GNLJVS pol polyprotein -
42 111 13.8 1101 1 B45390 pol polyprotein -
43 108.5 13.5 1087 2 J01162 pol protein - Maed
44 103 12.8 165 2 F84406 deoxycytidine triph
45 102 12.7 178 2 S26429 probable dCTP pyro

ALIGNMENTS

RESULT 1
E71216
dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: E71216
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hainaka, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137
A:Accession: E71216
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-156 <RAW>
A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31124.1; PID:g3258441
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH1997
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 95.3%; Score 768; DB 2; Length 156;
Best Local Similarity 93.6%; Pred. No. 1.le-64;
Matches 146; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
QY 1 MLLPDWKIRKEILIEPSEESLOPAGYDLRVGREAFAVKGKLDIVKRGKVVIPREVALI 60
Db 1 MLLPDWKIRKEILIEPSEESLOPAGYDLRVGREAFAVKGKLDIVKRGKVVIPREVALI 60
QY 61 LTLEKIKLPDVGMDKIRSLAREGVIGSAFWDPGWDGNLTIMLYNASNEPVELRYGE 120
Db 61 LTLEKIKLPDVGMDKIRSLAREGVIGSAFWDPGWDGNLTIMLYNASNEPVELRYGE 120
QY 121 RFVQIAPIRLEGFARNPYRGNVQSGTRLAFSKRKKL 156
Db 121 RFVQIAPIRLEGFARNPYRGNVQSGTRLAFSKRKKL 156

RESULT 2
G75030
dCTP deaminase (EC 3.5.4.13) PAB164 [similarity] - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75030
R:anonymous, Genoscope
A:Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: G75030

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB50685.1; PID:9545919
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: dcd; PAB1164
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 87.2%; Score 703; DB 2; Length 154;
Best Local Similarity 85.6%; Pred. No. 1.3e-58;
Matches 131; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLLPDWKIRKEILPEPSEESLOPAGYDLRVGREAFVKGKLDIVDEKGVVIPPVEYALI 60
Db 1 MLLPDWKIRKEILKPSEESLOPAGYDLRVGKEAYIQGKFIDVKEKGVLIIPPEYALI 60
QY 61 LTLERIKLPDDVMDKIRSLAREGVIGSFANVDPGWDGNLTMLYNASNEPVELRYGE 120
Db 61 LTLERIKLPDDINGMDKIRSLAREGVIGSFANVDPGWDGNLTMLYNASEKEVILRYKE 120
QY 121 RFVOIARIRLEGPARNPYRGNYQGSTRLAFSKR 153
Db 121 RFVOIAERLEAFKPNRYGNYQGSRRIVLSKR 153

RESULT 3
A70439
probable dCTP deaminase (EC 3.5.4.13) dcd [similarity] - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C:Accession: A70439
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ovi
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70439
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-180 <AQF>
A:Cross-references: GB:AE000747; NID:92983944; PIDN:AAC07499.1; PID:g2983951; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: dcd
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 30.1%; Score 243; DB 2; Length 180;
Best Local Similarity 37.7%; Pred. No. 1.4e-15;
Matches 61; Conservative 30; Mismatches 53; Indels 18; Gaps 5;

QY 1 MLLPDWKIRK-----EILIEPFSEESLOPAGYDLRVGRE-AFVKGK-LIDVEKGVV-- 51
Db 1 MILSDRSIRIELKELKVEPEFPHVQCSSLDLRLGNQIALYEGEGVIDVKRGTKGVRI 60
QY 52 -----IPPREVALITLTERIKLPDDVMDKIRSLAREGV-IGSFANVDPGWDGNL 102
Db 61 LEFEYEDIMPKQFLATTLEYISLPYVTAFAVEGRSSGLRLGFTIENAGWDAGFEQI 120
QY 103 TMLYNASNEPVELRYGERFVOIARIRLEGPARNPYRGNYQOG 144
Db 121 TLEFNANDRPIRLYGRMRCQLVFAFLDRPPEVYSGKYK 162

RESULT 4
H83695
deoxycytidine triphosphate deaminase BH0368 [Imported] - Bacillus halodurans (strain C-1
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000

C:Accession: H83695
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: H83695
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <STO>
A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA04087.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0368
C:Superfamily: dCTP deaminase

Query Match 30.1%; Score 242.5; DB 2; Length 177;
Best Local Similarity 37.2%; Pred. No. 1.6e-15;
Matches 58; Conservative 30; Mismatches 47; Indels 21; Gaps 6;

QY 10 KEILIEPFSEESLOPAGYDLRVGREAFV-----KGKLDIVKE-----EGKVVIP 54
Db 15 KELEITPLETEEQIPASVDLRLGPH-FVTIDDSKEAVISFERPIRYREWTSTDTIVLPP 73
QY 55 REYALITLTERIKLPDDVMDKIRSLAREGV-IGSFANVDPGWDGNLTMLYNASNEP 113
Db 74 HTFLLATTMETVKLPNHLTAFVEGRSSVGRGLFIQAGWVDFGNGQITLFLFNANRLP 133
QY 114 VELRYGERFVOIARIRLEGPARNPYRGNY---QGST 146
Db 134 IELPIGRRICQLVFAEVTGEVA-PYQGYLFQKGAT 168

RESULT 5
S26382
probable dCTP deaminase (EC 3.5.4.13) [similarity] - Desulfohalobus ambivalens
N:Alternate names: hypothetical protein 3 lig-region
C:Species: Desulfohalobus ambivalens
C:Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 31-Mar-2000
C:Accession: S26382
R:Kletzin, A.
Nucleic Acids Res. 20, 5389-5396, 1992
A:Title: Molecular characterization of a DNA ligase gene of the extremely thermophil
A:Reference number: S26382
A:Accession: S26382
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-173 <KLE>
A:Cross-references: EMBL:X63438; NID:g40784; PIDN:CA45033.1; PID:g40785
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1991
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 30.0%; Score 242; DB 2; Length 173;
Best Local Similarity 35.9%; Pred. No. 1.7e-15;
Matches 56; Conservative 30; Mismatches 56; Indels 14; Gaps 3;

QY 12 ILIEPFSEESLOPAGYDLRVG-----REAFVKGK-----LIDVEKGVVIPPVEYA 58
Db 17 IVISPLTQDTIRENGVDLVRVGGEIARFKTDEIVEDGKDPSEFVEIEKGDEFIYPNEHV 76
QY 59 LIITLTERIKLPDDVMDKIRSLAREGVIGSFANVDPGWDGNLTMLYNASNEPVELRY 118
Db 77 LLVTEEVKLPNDVMAFVNLRSFARLGLFVPPITVDAGFEGQITIEVL-GSAPPVKIKR 135
QY 119 GERFVOIARIRLEGPARNPYRGNYQGSTRLAFSKR 154
Db 136 GTRFLHLIFARTLTPVENPYHGKYGQGGVTLPKFK 171

RESULT 6
D27274

probable dCTP deaminase (EC 3.5.4.13) APE0333 [similarity] - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: D72724
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339
A:Accession: D72724
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-181 <RAW>
A:Cross-references: DDBJ:AF000059; NID:g5103911; PIDN:BAA79288.1; PID:g5103972
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0333
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 28.8%; Score 232.5; DB 2; Length 181;
Best Local Similarity 31.8%; Pred. No. 1.4e-14;
Matches 55; Conservative 38; Mismatches 57; Indels 23; Gaps 4;
QY 1 MLLPDMKIRK-----EILIEPFSEESLQAGYDLRVGREAFAVK-----G 39
Db 4 LILSDRDIRALLAIGDLIVPEPLSGDVTRENGLDLRLGR-GCFRKRSDRVLDPRAPGSPG 62
QY 40 KLIDVEKEGVVIPPYREALILTLERIKLPDDVMDKIRSLAREGVISGFAWVDPGDW 99
Db 63 EFYECGGEDIIIVGPGHMLLHTQEIYRLPGYVAGLVNLRSTWARTGIYIPATVVDAGFE 122
QY 100 GNLTMLYNASNPVELRYGERVQIAFIRLEGPARNPYRGNGVGGSTRFAFSK 152
Db 123 GQLTIEVV-GSGPVPKLYPDGRELHLVLVQLQSPAMNYPYRGYQGGVRLPK 174

RESULT 7
T36613
probable dCTP deaminase (EC 3.5.4.13) SCH35.46 [similarity] - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: T36613
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, June 1999
A:Reference number: Z21610
A:Accession: T36613
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-191 <OLI>
A:Cross-references: EMBL:AL078610; PIDN:CAB44381.1; GSPDB:GNO0070; SCOEDB:SCH35.46
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCH35.46
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 26.8%; Score 216; DB 2; Length 191;
Best Local Similarity 31.4%; Pred. No. 5.2e-13;
Matches 53; Conservative 32; Mismatches 60; Indels 24; Gaps 3;
QY 1 MLLPDMKIRKEI-----LIEPFSEESLQAGYDLRVGR-----EAFVK 38
Db 1 MLLSDKDIRAIDNGRVIRDPDMSVMQPSIDVRLDRYFRVFENHRYPHIDPSVQVDL 60
QY 39 GKLIIDVEKEGVVIPPYREALILTLERIKLPDDVMDKIRSLAREGVIT--GSFAWVDP 96
Db 61 TLRIVEEGDPEFLIHGGEFVLAITYEWSLPDDLASRLGKSSGLRLGLIYTHSTAGFIDP 120
QY 97 GWDGNLTMLYNASNPVELRYGERVQIAFIRLEGPARNPYRGNGVGS 145

Db 121 GFSCHVTLESLNLTLPKLPQMGKIQGLCLFRLTSPAEPYGSERYGS 169

RESULT 8
A69114
dCTP deaminase (EC 3.5.4.13) MTH1847 [similarity] - Methanobacterium thermoautotrophophilum
C:Species: Methanobacterium thermoautotrophophilum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 31-Mar-2000
C:Accession: A69114
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwanji, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N. J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophophilum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: A69114
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-200 <MTH>
A:Cross-references: GB:AE000937; GB:AE000666; NID:g2622974; PIDN:AAB86313.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1847
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 24.7%; Score 199; DB 2; Length 200;
Best Local Similarity 31.6%; Pred. No. 2.1e-11;
Matches 59; Conservative 29; Mismatches 63; Indels 36; Gaps 7;
QY 2 LLLPDMKIRKE-----LIEPFSEESLQAGYDLRVGREAFAVKGLI----- 42
Db 6 ILSDRLKRIYIEGLITIDLPDPERIQSPSSVDLRIGNE--FKGFRVIRKPCIDPKDPS 63
QY 43 DVE-----KEGVVIPPYREALILTLERIKLPDDVMDKIRSLAREGVIT--GSFA 92
Db 64 DIESYMETPHVEDGPFTHPGEFALATTHEYIALPEDLVARVEGRSSIGRLGITMHTAG 123
QY 93 WVDPGWGNLTMLYNASNPVELRYGERVQIAFIRLEGPARNPY-----RGNVQSGSTR 147
Db 124 YIDPGFHRITLISNIGKMPVALYPRQVCQIVFETMTSPAERPYPHPSRDSKYIGQTR 183

QY 148 LAFSKRK 154
Db 184 PQTSTRK 190

RESULT 9
B70526
dCTP deaminase (EC 3.5.4.13) RV0321 [similarity] - Mycobacterium tuberculosis (strain C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70526
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: B70526
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-190 <COL>
A:Cross-references: GB:Z96800; GB:AL123456; NID:g3261800; PIDN:CAB09605.1; PID:g21939
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: dcd
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 24.6%; Score 198; DB 2; Length 190;
Best Local Similarity 30.2%; Pred. No. 2.5e-11;
Matches 51; Conservative 33; Mismatches 61; Indels 24; Gaps 3;
C:Superfamily: dCTP deaminase

QY 1 MLLPDMKIRKIL-----IIPFSESLQAGYDLRV-----GREAFVK 38
DB 1 MLLSDRLRAEISSGRIGIDFDLTVQPSIDVLOCLRFVNNTRYTHIDPAKQODEL 60
QY 39 KGLIDVEKEGVVIPPPEYALILTLERIKLPDDVGMKIRSSLAREGVV--GSFAWVDP 96
DB 61 TSLVQPVGDGEPVLPQGEFVLGSLTEFLTPDNLNAGLEKSSLRGLGUTHSTAGFIDP 120
QY 97 GWDGNLTMLYNASNEPVELRGVRFVQIAFIRLEGPARNPYRGNYQS 145
DB 121 GFSGHITLSELSNVANLITLWPKMGKIQGLCMRLTSPSEHPYGSRRAGS 169
RESULT 10
A81272
probable dCTP deaminase (EC 3.5.4.13) Cj1292 [imported] - Campylobacter jejuni (strain N
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000
C:Accession: A81272
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912
A:Accession: A81272
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-186 <PAR>
A:Cross-references: GB:AL139078; GB:AL111168; NID:96968723; PIDN:CA873719.1; PID:9696872
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: dcd; Cj1292
A:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 23.5%; Score 189.5; DB 2; Length 186;
Best Local Similarity 29.4%; Pred. No. 1.5e-10;
Matches 53; Conservative 34; Mismatches 58; Indels 35; Gaps 7;
C:Superfamily: dCTP deaminase

QY 5 DWKIRKIL-----IIPFSESLQAGYDLRV-----AF 36
DB 7 NW-IRKMALEHKMIPEPCANIGKGVVSYGLSSYGYDIRVGRFKIFTNVNSTVDPKPF 65
QY 37 VKGKLIDVEKEGVVIPPPEYALILTLERIKLPDDVGMKIRSSLAREGVVGSFAWVD 95
DB 66 VEENVVDF--EGDVCIYPANSAFALARTIEYFKMPDNVLAICLGSYARCGIIVNTPFE 123
QY 96 PGWDGNLTMLYNASNEPVELRGVRFVQIAFIRLEGPARNPY---RGNTQGSTRLAFSK 152
DB 124 PGFEGHTIETNSPTPLPAKIYANEGIAVQLFLOGDEKCDTTRYKDKKGYQAQGTITLPR 183
RESULT 11
G84184
deoxycytidine triphosphate deaminase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: G84184
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; L
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: G84184
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-195 <STO>
A:Cross-references: GB:AE004437; NID:g10579885; PIDN:AAG18843.1; GSPDB:GN00138
C:Genetics:
A:Gene: dtd
C:Superfamily: dCTP deaminase

Query Match 22.9%; Score 184.5; DB 2; Length 195;
Best Local Similarity 31.2%; Pred. No. 4.6e-10;
Matches 54; Conservative 31; Mismatches 59; Indels 29; Gaps 7;
C:Superfamily: dCTP deaminase

QY 1 MLLPDMKIRK-----EILIPFSESL--QPAGYDLRVGRE--AFVKGKL-----41
DB 1 MILSDQDILARLADGDLAIEPLEVDVLQVQSPASVDVLRGRLEFEFERNVPCIHNPNEDE 60
QY 42 ID-----VEKEGVVIPPPEYALILTLERIKLPDDVGMKIRSSLAREGVV--GSFA 92
DB 61 VDEYVTVETVVEDGDEFILHPGDFVLGTTKERVEVPRDLVAQVEGRSSILGLAVVVHATAG 120
QY 93 WVPDGMGNLTMLYNASNEPVELRGVRFVQIAFIRLEGPARNPYRGNYQS 145
DB 121 FIDEGFNGRYTLELSNLGKVPVALTPEMRISQLVFTELTSPADRPY-GDERGS 172
RESULT 12
D64566
dCTP deaminase (EC 3.5.4.13) HP0372 [similarity] - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Mar-2000
C:Accession: D64566
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; Mcke
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: D64566
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-190 <TOM>
A:Cross-references: GB:AE000554; GB:AE000511; NID:g2313475; PIDN:AAD07441.1; PID:g231
C:Genetics:
A:Start codon: GTG
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 22.5%; Score 181; DB 1; Length 190;
Best Local Similarity 28.0%; Pred. No. 9.5e-10;
Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;
C:Superfamily: dCTP deaminase

QY 13 LIIPFSEE-----SLQAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
DB 20 MISPFCEKQKGVNVSIGLSSYGYDIRVGRSEFMLFDKNALIDPKNFDPNNATKIDASKE 79
QY 48 GKVVIPPPEYALILTLERIKLPDDVGMKIRSSLAREGVVGSFAWVDGNGNLTMLY 107
DB 80 GYFILPANAFALARTIEYFKMPKDTLAICLGSYARCGIIVNTPPEFEGYITIEIS 139
QY 108 NASNEPVELRGVRFVQIAFIRLEGPARNPY---GNVQGSTRLAFSK 152
DB 140 NTNLPAKVYANEGIAVQVFLQGDCEMQSYKDRGKGYQGVGTTLPK 187
RESULT 13
E71860
dCTP deaminase (EC 3.5.4.13) dcd [similarity] - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Mar-2000
C:Accession: E71860
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: E71860
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <ARN>
A:Cross-references: GB:AE001529; GB:AE001439; NID:q4155590; PIDN:AAD06585.1; PID:g415559
A:Experimental source: strain J99
C:Genetics:
A:Gene: dcd
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 22.38; Score 180; DB 2; Length 188;
Best Local Similarity 28.08; Pred. No. 1.2e-09;
Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;
Qy 13 LIEPFSE-----SLOPAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
Db 18 MINFCEKQVKNVISGLSYGDIRVGFEMFLDNKNALIDPKNFPNNAKIDASKE 77
Qy 48 GKVVIPPREYALILTLEKIPDDVMDKIRSSLAKEGVIGSFAWVDPGWDGNLTMLY 107
Db 78 GFFILPANAFALAHTEYFKMPKDTLAICLGKSTYARGCIIVNVTPEPEGVITIBIS 137
Qy 108 NASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNVQSTRLAFSK 152
Db 138 NTTNLPKAVYANEGVIAQVVFLOQGEVCEQSKYKDRGGKYQGQGVITLPK 185

RESULT 14
C84942
dCTP deaminase (EC 3.5.4.13) [imported] - Buchnera sp. (strain APS)
N:Alternate names: deoxycytidine triphosphate deaminase
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A:Reference number: A84930; MUID:20445173
A:Accession: C84942
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: dcd; BU108
C:Keywords: hydrolase

Query Match 21.68; Score 174; DB 2; Length 206;
Best Local Similarity 29.38; Pred. No. 4.7e-09;
Matches 48; Conservative 30; Mismatches 58; Indels 28; Gaps 5;
Qy 3 LPDWKIRKEILIEPFSESL-OPAGYDLRVGRE--AFVK--GKLIDV----- 44
Db 21 IEWLERKELIIEPYPNKTLINGITVDIHLGNKFRFFVEHTGSCIDLSNKSIIIGLSITE 80
Qy 45 -----EKEGKVIPPREYALILTLEKIPDDVMDKIRSSLAKEGV---GSPAW 94
Db 81 IMSNEIIFSKQPCFLQPGSLVLCSTFESIKMNNPWLGDGRSLARLGLMIHATAHRI 140
Qy 95 DPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY 138
Db 141 DPGWNGNIVLEMFNAGKLTIVLRPKMRIALUSFEVLSPQLRYP 184

RESULT 15

A83210
probable deoxycytidine triphosphate deaminase PA3480 [imported] - Pseudomonas aerugin
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83210
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: A83210
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-188 <STO>
A:Cross-references: GB:AE004769; GB:AE004091; NID:g9949624; PIDN:ARG06868.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3480
C:Superfamily: dCTP deaminase

Query Match 21.38; Score 172; DB 2; Length 188;
Best Local Similarity 26.38; Pred. No. 6.5e-09;
Matches 45; Conservative 32; Mismatches 62; Indels 32; Gaps 4;
Qy 13 LIEPFSESLQPA-----GYDLRVGRE-----AFVKGKLDIV 44
Db 18 MIEPFVERQVRGADDSRVISYGVSSYGYDVRCAAEFKVTNIHSAVVDPKNFDEKSFVDI 77
Qy 45 EKEGKVIPPREYALILTLEKIPDDVMDKIRSSLAKEGVIGSFAWVDPGWDGNLT 104
Db 78 NSD-VCIIPNPSFALARTVEYFRIPRODLTICLGKSTYARGCIIVNVTPEPEGHVTL 136
Qy 105 MLYNASNEPVELRYGERFVQIAFIRLEGPARNPYR---GNVQSTRLAFSK 152
Db 137 EFSNTNLPKAVYANEGVIAQVVFLOQGEVCEVSKYKDRGGKYQGQGVITLPK 187

Search completed: January 31, 2002, 13:08:39
Job time: 46 sec

This Page Blank (uspto)

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	768	95.3	156	1	DCD_PYRHO	Q57706	pyrococcus
2	703	87.2	154	1	DCD_PYRAB	Q9uxs8	pyrococcus
3	257.5	31.9	183	1	DCD_THEAC	Q9hk00	thermoplasma
4	243	30.1	180	1	DCD_AQAE	Q67539	aquifex aerophilus
5	242	30.0	173	1	DCD_ACIAM	Q02103	acidianus acidophilus
6	232.5	28.8	181	1	DCD_AERPE	Q9y8a8	aeropyrum pyroaerophilum
7	216	26.8	191	1	DCD_STRCO	Q9x8w0	streptomyces coelicolor
8	199	24.7	197	1	DCD_METTH	Q27875	methanobacterium thermoautotrophicum
9	198	24.6	190	1	DCD_MYCTH	Q07247	mycobacterium tuberculosis
10	181	22.5	188	1	DCD_HELPJ	Q25136	helicobacter pylori
11	180	22.3	188	1	DCD_HELPJ	Q9xkd0	helicobacter pylori
12	174	21.6	193	1	DCD_BUCAI	P57209	buchnera aphidivorus
13	163	20.2	193	1	DCD_ECOLI	P28248	escherichia coli
14	161.5	20.0	193	1	DCD_BUCAP	Q9zhd8	buchnera aphidivorus
15	154	19.1	204	1	DCD_METJA	Q57872	methanococcus jannaschii
16	150.5	18.7	190	1	DCD_CHLMU	Q9pkz9	chlamydia muridarum
17	149.5	18.5	195	1	DCD_CHLTR	Q84042	chlamydia trachomatis
18	149.5	18.5	195	1	DCD_HAEIN	P44534	haemophilus influenzae
19	146.5	18.2	194	1	DCD_PASMO	P57891	pasteurella multocida
20	145	18.0	188	1	DCD_RICPR	Q9ze77	rickettsia prowazekii
21	138	17.1	190	1	DCD_CHLNP	Q9z8f1	chlamydia pneumoniae
22	137	17.0	172	1	DCD_GLOHI	Q9znj8	clostridium histolyticum
23	125.5	15.6	1145	1	POL_ETAVY	Q03371	equine influenza virus
24	125.5	15.6	1146	1	POL_ETAV9	P11204	equine influenza virus
25	125.5	15.6	1146	1	POL_ETAVC	P32542	equine influenza virus
26	118	14.6	1109	1	POL_CAEVC	P33459	caprine arthritis-encephalitis virus
27	111.5	13.8	1086	1	POL_OMVVS	P16901	ovine lentivirus
28	111	13.8	252	1	DUT_HUMAN	P33316	homo sapiens immunodeficiency virus
29	111	13.8	1101	1	POL_VILVK	P35956	viena lentivirus
30	111	13.8	1105	1	POL_VILVI	P03370	viena lentivirus
31	111	13.8	1105	1	POL_VILV1	P23426	viena lentivirus
32	111	13.8	1105	1	POL_VILV2	P23427	viena lentivirus
33	105.5	13.1	140	1	DUT_SCHPS	Q9p6q5	schistosoma mansoni

Db 61 LTLERKLPDDVMDGKIRSSLAGREGVLGSPAWDPGWDGNTLMLYNASNEPVELRYKE 120
Qy 121 RFVQIAFLRLEGPARNPYRGNYQGSTRLAFSKRKL 156
Db 121 RFVQIAFLRLEGPARNPYRGNYQGSTRLAFSKRKL 156

RESULT 2
ID DCD PYRAB STANDARD; PRT; 154 AA.
AC Q9UXSR:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE)
GN DCD OR PA81164
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=49292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Helling R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248288; CAB50685.1;
DR InterPro; IPR003232; dCTP_deaminase.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 154 AA; 17758 MW; C031BEE419094DDB CRC64;

Query Match 87.28; Score 703; DB 1; Length 154;
Best Local Similarity 85.66; Pred. No. 7e-60; Mismatches 9; Indels 0; Gaps 0;
Matches 131; Conservative 13;
Qy 1 MLLPDWKIRKEILIEPSEESLQAGYDLRVGREAFAVKGKLDIVKEKGVVIPPYVALI 60
Db 1 MLLPDWKIRKEILIEPSEESLQAGYDLRVGREAFAVKGKLDIVKEKGVVIPPYVALI 60

Qy 61 LTLERKLPDDVMDGKIRSSLAGREGVLGSPAWDPGWDGNTLMLYNASNEPVELRYKE 120
Db 61 LTLERKLPDDVMDGKIRSSLAGREGVLGSPAWDPGWDGNTLMLYNASEKEVILRYKE 120

Qy 121 RFVQIAFLRLEGPARNPYRGNYQGSTRLAFSKR 153
Db 121 RFVQIAFLRLEGPARNPYRGNYQGSTRLAFSKR 153

RESULT 3
ID DCD THEAC STANDARD; PRT; 183 AA.
AC Q9HKK0:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE).
GN DCD OR TA0598.

OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermophilic scavenger Thermoplasma
RT acidophilum";
RL Nature 407:508-513(2000)
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL445064; CAC11737.1;
DR InterPro; IPR003232; dCTP_deaminase.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 183 AA; 20245 MW; 41D1992A99CD6682 CRC64;

Query Match 31.98; Score 257.5; DB 1; Length 183;
Best Local Similarity 41.38; Pred. No. 1.5e-17;
Matches 64; Conservative 22; Mismatches 50; Indels 19; Gaps 5;
Qy 1 MLLPDWKIRKEILIEPSEESLQAGYDLRVGREAFAVKGKLDIVKEKGVVIPP 53
Db 18 MILNDSTIMRMVSDGLLISENDFRGCLTPNGYDLRV-----DAIDVEGRQYSEFEG 69

Qy 54 PREVALILTLEIKLPDDVMDGKIRSSLAGREGVLGSPAWDPGWDGNTLMLYNASNEP 113
Db 70 KNVHFLVSTTEILKIPDDVMDGKIRSSLAGREGVLGSPAWDPGWDGNTLMLYNASNEP 128

Qy 114 VELRYGERFVQIAFLRLEGPARNPYR---GNVYGS 145
Db 129 VNLRGRIQIVFVKMIGSAEKPYHRSNGYQNS 163

RESULT 4
ID DCD AQUAE STANDARD; PRT; 180 AA.
AC O67539;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR AQ.1607.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujaay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).

CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000747; AAC07499.1; -
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 180 AA; 20544 MW; B2710421A2FA48D6 CRC64;

Query Match 30.1%; Score 243; DB 1; Length 180;
Best Local Similarity 37.7%; Pred. No. 3.5e-16;
Matches 61; Conservative 30; Mismatches 53; Indels 18; Gaps 5;

Qy 1 MLLPDWKIRK-----EHLIEPFSEESLQAGYDLRVG--AFVKGK-LIDVEKEGVV-- 51
Db 1 MILSDRSIRELIEKGELKVEPEYSPHVCSSLDRLGNQIALYEGEGVIDVKKTKGVRI 60
Qy 52 -----IPREVALILTLEIRIKLPDDVGMCKIRSLAREGV-IGSAFWDPGWDGNL 102
Db 61 LEFEYFDIMPKQFLATTLEYISLPYVTAFFVGRSLGRLGFIENAGWVDAFGQI 120
Qy 103 TLMLYNASNEVELRGERFVOIAFIRLEGPARNPYRGNYQG 144
Db 121 TLELFANDRPIRYLGRMRCQLVFLDRPPERYSGKYKG 162

RESULT 5
DCD_ACIAM STANDARD; PRT; 173 AA.
AC Q02103:
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD.
OS Acidianus ambivalens (Desulfurolobus ambivalens).
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Acidianus.
OX NCBI_TaxID=2283;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Le1 10 / DSM 3772;
RX MEDLINE=93065206; PubMed=1437556;
RA Kletzin A.;
RT "Molecular characterisation of a DNA ligase gene of the extremely
RT thermophilic archaeon Desulfurolobus ambivalens shows close
RT phylogenetic relationship to eukaryotic ligases.";
RL Nucleic Acids Res. 20:5389-5396(1992).
RN [2]
RP SIMILARITY.
RX MEDLINE=95206934; PubMed=7899076;
RA Ouzounis C., Kyriades N., Sander C.;
RT "Novel protein families in archaean genomes.";
RL Nucleic Acids Res. 23:565-570(1995).
CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X63438; CAA45033.1; -
DR PIR: S26382; S26382.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hydrolase.
SQ SEQUENCE 173 AA; 19858 MW; B4D922503CD4B25A CRC64;

Query Match 30.0%; Score 242; DB 1; Length 173;
Best Local Similarity 35.9%; Pred. No. 4.1e-16;
Matches 56; Conservative 30; Mismatches 56; Indels 14; Gaps 3;

Qy 12 ILIEPFSEESLQAGYDLRVG-----REAFVKGK----LIDVEKEGVVPPREYA 58
Db 17 IVISPLTQDITRENGVDLRVGGEIARFKKTDIEYEDGKDPKRSFYIEKGDEFIYPNEHV 76
Qy 59 LILFLERIKLPDDVGMCKIRSLAREGVIGSAFWDPGWDGNLTLMLYNASNEVELRY 118
Db 77 LLVTEEVYKLPNDVMAFVNLRSSFARLGLFVPPITVDAGFEGQLTIEVL-GSAFPVKIKR 135
Qy 119 GERFVOIAFIRLEGPARNPYRGNYQGSTRLAFSKRK 154
Db 136 GTRFLHLFIARTLTVPENPYHGKYGQGGVTLPKFK 171

RESULT 6
DCD_AERPE STANDARD; PRT; 181 AA.
AC Q5VF38:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR APE0333.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP000059; BAA79288.1; -
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.

```
KW Hydrolase; Complete proteome.
SQ SEQUENCE 181 AA; 19894 MW; D8B6CBDC1722EFE9 CRC64;

Query Match 28.8%; Score 232.5; DB 1; Length 181;
Best Local Similarity 31.8%; Pred. No. 3.5e-15;
Matches 55; Conservative 38; Mismatches 57; Indels 23; Gaps 4;

QY 1 MLLPDWKIRK-----ELIEPFSESLOPAGYDLRVGREAFVK-----G 39
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 4 LILSDRDIRALLAIGDVLVEPLSGDTVRENGDLRLGR-GFCRFKRSRDVLDPRAPGSPG 62
   :| | | | | :| | | | | :| | | | | :| | | | |

QY 40 KLIDVEKEGVVIPPPEYALITLERIKLPDDVGMKIRSSLAAREGVIGSFVWDPGWD 99
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 63 EFVECGEGDEIIIVGPGHMLLTQEIYRLPGYVAGVNLRSWTARTGIYIPATVVVDAGFE 122
   :| | | | | :| | | | | :| | | | | :| | | | |

QY 100 GNTLTMLYNASNEPVELRYGERFVOIAFIRLEGPARNPYRGNGVSTRLAFSK 152
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 123 GOLTIEVV-GSGFPVKVLPQDRFLHLVLKLOSPAMPNRYGRYQGRGVRLPK 174
   :| | | | | :| | | | | :| | | | | :| | | | |

RESULT 7
DCD_STRCO
ID DCD_STRCO STANDARD; PRT; 191 AA.
AC QX8W0; 2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR SCH35.46.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,
RA Rajadream M.A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AL078610; CAB44381.1; -
DR InterPro: IPR003232; dCTP_deaminase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hydrolase.
SQ SEQUENCE 191 AA; 21496 MW; 6352A496990F910C CRC64;

Query Match 26.8%; Score 216; DB 1; Length 191;
Best Local Similarity 31.4%; Pred. No. 1.4e-13;
Matches 53; Conservative 32; Mismatches 60; Indels 24; Gaps 3;

QY 1 MLLPDWKIRKEI-----LIEPFSESLOPAGYDLRVGR-----EAFVK 38
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 1 MLLSKDIRAIEDGRVIDPFDDSMVQPSIDVRLDYRFVFNHRYPHIDPSVEQVDL 60
   :| | | | | :| | | | | :| | | | | :| | | | |

QY 39 GKLDVKEKEGVVIPPPEYALITLERIKLPDDVGMKIRSSLAAREGVI--GSFAWDP 96
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 61 TRLVEPGDEPFILHNGEFVLASTYEVVSLPDDLASLRGKSSLRGLVTHSTAGFIDP 120
   :| | | | | :| | | | | :| | | | | :| | | | |

KW Hydrolase; Complete proteome.
SQ SEQUENCE 181 AA; 19894 MW; D8B6CBDC1722EFE9 CRC64;

Query Match 24.7%; Score 199; DB 1; Length 197;
Best Local Similarity 31.6%; Pred. No. 5.7e-12;
Matches 59; Conservative 29; Mismatches 63; Indels 36; Gaps 7;

QY 2 LLPDWKIRKE-----ILIEPFSESLOPAGYDLRVGREAFVKGLI----- 42
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 3 ILSDRDIRKRYIEEGTLITIDLPDPERQIQSPSSVDLRICNE--FKGFRVIRKPCIDPKDPS 60
   :| | | | | :| | | | | :| | | | | :| | | | |

QY 43 DVE-----KEGVVIPPPEYALITLERIKLPDDVGMKIRSSLAAREGVI--GSFA 92
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 61 DIESMETFHEVDGPFIIHPGFEFALATTHEYIALPEDLVARVEGRSSIGRLGITMHTAG 120
   :| | | | | :| | | | | :| | | | | :| | | | |

QY 93 WVPDGDGNLTMLYNASNEPVELRYGERFVOIAFIRLEGPARNPY-----RGNVGGSTR 147
   :| | | | | :| | | | | :| | | | | :| | | | |
Db 121 YIDPGFGRITLISNIGKMPVALYPRQVCQIVFETMTSPAERPYPCHPSRDSKYIGQTR 180
   :| | | | | :| | | | | :| | | | | :| | | | |

QY 148 LAFSKR 154
Db 181 PQTSRIK 187
```

```

Query Match      24.6%; Score 198; DB 1; Length 190;
Best Local Similarity 30.2%; Pred. No. 6.9e-12;
Matches 51; Conservative 33; Mismatches 61; Indels 24; Gaps 3;

QY 1 MLLPDKWKRIEIL-----IEPSEESLQAPGYDLRV-----GREAFAVK 38
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MLLSDRLRAEITSSGRLGIDPFDTLVLQVFSNIDVRLDCLFRVFNNTRYTHIDPAKQODEL 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```
RT dut (dUTPase) mutations." ;
RL J. Bacteriol. 174:5647-5653(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "the complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map";
RL DNA Res. 3:379-392(1996).
CC -1- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -1- SUBUNIT: HOMOTETRAMER (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M90069; AAA23669.1; -.
DR EMBL; AE000296; AAC75126.1; -.
DR EMBL; D90844; BAA15918.1; -.
DR EMBL; D90845; BAA15923.1; -.
DR F01; A42940; A42940.
DR EMBL; EG11418; dcd.
DR InterPro; IPR003232; dCTP_deaminase.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 193 AA; 21249 MW; B0044051ADE7F919 CRC64;

Query Match 20.2%; Score 163; DB 1; Length 193;
Best Local Similarity 29.9%; Pred. No. 1.5e-08;
Matches 50; Conservative 23; Mismatches 60; Indels 34; Gaps 5;

QY 20 ESQAPAGYDLRVGRE-----AFVK-----GKLIDVEKEGVIP-----P 54
DB 26 ERINGATVDYRLGNKFTFRGHFAAFIDLSGPKDEVSALDRVMSDEIVLDEGEAFVLP 85
QY 55 REVALIITLERIKLPDDVMGDMKIRSSLAEGVIGSPA--WVDPGWGNTLTMLYNASNE 112
DB 86 GELALAVTLESVTLPADLVGWLGRSSRLARGLMVHVHTAHRIDPGWGGCIVLEFYNSGKL 145
QY 113 PVELRYCERFVOTAFIRLEGPARNPYR-----GNVQSGTRIAFSK 152
DB 146 PLALRPGMLGIALSFEPLSGPAVRPNRREDAKYRNQOGGAVASRIDK 192

RESULT 14
DCTP_BUCAP
ID DCD_BUCAP STANDARD; PRT; 193 AA.
AC Q9ZHD8;
DT 20-AUG-2001 (Rel. 40, Created)

dnt (dUTPase) mutations." ;
RL J. Bacteriol. 174:5647-5653(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "the complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=98440331; PubMed=9767718;
Clark M.A., Baumann L., Baumann P.;
RT "Buchnera aphidicola (Aphid endosymbiont) contains genes encoding
enzymes of histidine biosynthesis.";
RL Curr. Microbiol. 37:356-358(1998).
CC -1- CATALYTIC ACTIVITY: dCTP + H(2)O = dUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF067228; AAC97363.1; -.
DR InterPro; IPR003232; dCTP_deaminase.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
KW Hydrolase.
SQ SEQUENCE 193 AA; 21871 MW; 8E19D4580C7C55E8 CRC64;

Query Match 20.0%; Score 161.5; DB 1; Length 193;
Best Local Similarity 26.0%; Pred. No. 2e-08;
Matches 47; Conservative 33; Mismatches 66; Indels 35; Gaps 5;

QY 3 LPDWKIRKEILIEPFSESL-QPAGYDLRVGRE-----AFVKGK 40
DB 8 IEELWSKKLVIQIOPYPKKLINGITVDIHGNKFRFFDYHTSCIDLSGSKKIALDLNK 67
QY 41 LIDVE-----KEGVVIPPVEYALITLERIKLPDDVMGDMKIRSSLAEGVIGSPA--WV 94
DB 68 IVSCETIFSKKEPFELKPGALALEFSTLENTLPNNLVGWLGRSSRLARGLMVHVHTSHRI 127
QY 95 DPWGDNLTMLYNASNEPVELRYCERFVOTAFIRLEGPARNPYRGNVQSGTRIAFSK 154
DB 128 DPWGHGNIIVLEFFNAGKLTDLVTPGKIAALSFEILSKPVLRLPYNSRNE-----SKYK 180
QY 155 K 155
DB 181 R 181

RESULT 15
DCTP_METJA
ID DCD_METJA STANDARD; PRT; 204 AA.
AC Q57872;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
DE DCD OR MJO430.
OS Methanococcus jannaschii.
OS Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus
OC NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
```

```

RX MEDLINE=96337999; PubMed=688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kleravag A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.D., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: U67494; AAB98415.1; -
DR TIGR: M70430; -
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001426; dUTPase.
DR Pfam: PF00562; dUTPase.1.
DR ProDom: PD004900; dCTP_deaminase; 1.
DR KEGG: Complete proteome.
SQ SEQUENCE 204 AA; 23432 MW; 1218368057723371 CRC64;

```

Query Match 19.18; Score 154; DB 1; Length 204;
Best Local Similarity 25.78; Pred. No. 1.le-07;
Matches 47; Conservative 37; Mismatches 61; Indels 38; Gaps 6;

```

Qy 5 DWKIRKEILIEPFSEESLQPGYDLRVGREAFV-KGKLIDVEKE-----GKVVIPP 54
Db 10 DYVTSKRRIIKPFKNKDFVPCSYDVTILGDEFIYDDEYDLSKELNYKRIKIKNSILVCP 69
Qy 55 REYALI-----LTLEIRIKLPDDVYMGDMKIRSSLAEGVIG-- 89
Db 70 LNYNLTEEKINYPKEKYNVDYVVEGGVLGTTNEYIELPNDISAOYQGRSSLGVRVFLTSHQ 129
Qy 90 SFANVDPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNYOGSTRLA 149
Db 130 TAGWIDAGFGKGTLEIV-AFDKPVILYKNORIGQLIFSKLLSPADVGY--SERKTSKYA 186
Qy 150 FSK 152
Db 187 YQK 189

```

Search completed: January 31, 2002, 13:09:23
Job time: 39 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:13:34 ; Search time 35.08 seconds
(without alignments)
650.470 Million cell updates/sec

Title: US-08-957-709-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKEILIEPFSEE.....PYRGNQSGSTRLAFSKRKL 156

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_17.*
- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	242.5	30.1	177	2 Q9KFV3	Q9KFV3 bacillus ha
2	237	29.4	158	12 O71028	O71028 archaeal vi
3	197.5	24.5	190	2 Q9CB17	Q9CB17 mycobacteri
4	189.5	23.5	186	2 Q9PN07	Q9PN07 campylobact
5	184.5	22.9	195	1 Q9HSG3	Q9HSG3 halobacteri
6	172	21.3	188	2 Q9HFC9	Q9HFC9 pseudomonas
7	164.5	20.4	191	2 Q9PFB6	Q9PFB6 xylella fas
8	159	19.7	163	1 Q9YG32	Q9YG32 aeropyrum p
9	155	19.2	188	2 Q9JRE8	Q9JRE8 neisseria m
10	153	19.0	150	1 O27642	O27642 methanobact
11	135.5	16.8	193	2 P74073	P74073 synechocyst
12	126.5	15.7	1146	12 Q9B6G1	Q9B6G1 equine infe
13	126	15.6	168	1 O29157	O29157 archaeoglob
14	125.5	15.6	170	12 Q9WHF3	Q9WHF3 agrotis seg
15	125.5	15.6	1138	12 Q9W7U6	Q9W7U6 equine infe
16	125.5	15.6	1146	12 O89468	O89468 equine infe
17	125.5	15.6	1146	12 O89472	O89472 equine infe
18	123.5	15.3	422	12 Q66733	Q66733 equine infe
19	123.5	15.3	1148	12 Q9EP46	Q9EP46 equine infe

20	123	15.3	199	2	Q9RMP1	Q9RMP1 zymomonas m
21	120.5	15.0	1134	12	Q992J6	Q992J6 equine infe
22	119.5	14.8	1148	12	Q9EP41	Q9EP41 equine infe
23	117.5	14.6	161	1	Q58502	Q58502 methanococc
24	117.5	14.6	1134	12	Q992K2	Q992K2 equine infe
25	116.5	14.5	1107	12	Q9DKV8	Q9DKV8 caprine art
26	115.5	14.3	188	5	Q9V311	Q9V311 drosophila
27	103	12.8	165	1	Q9HMF3	Q9HMF3 halobacteri
28	102	12.7	148	2	Q9AIK2	Q9AIK2 streptococc
29	101	12.5	162	11	Q9JJ44	Q9JJ44 mus musculu
30	101	12.5	162	11	Q9CQ43	Q9CQ43 mus musculu
31	101	12.5	204	11	Q9CU90	Q9CU90 mus musculu
32	100.5	12.5	160	12	O72165	O72165 orf virus.
33	99	12.3	1086	12	Q84809	Q84809 puma lentiv
34	98.5	12.2	166	10	Q9STG6	Q9STG6 arabidopsis
35	96.5	12.0	430	12	Q9WMY7	Q9WMY7 dioscorea a
36	96.5	12.0	430	12	Q98W23	Q98W23 dioscorea a
37	93.5	11.6	145	12	Q9J5G5	Q9J5G5 fowlpox vir
38	93	11.5	143	12	Q9J880	Q9J880 spodoptera
39	90.5	11.2	148	12	Q9Q8S9	Q9Q8S9 myxoma viru
40	89.5	11.1	147	12	P87630	P87630 cowpox viru
41	89.5	11.1	147	12	Q9JFF1	Q9JFF1 vaccinia vi
42	87.5	10.9	1123	12	O66933	O66933 feline immu
43	87.5	10.9	1150	12	P90246	P90246 feline immu
44	87	10.8	155	2	Q9A253	Q9A253 caulobacter
45	86.5	10.7	317	12	O10287	O10287 orgyia pseu

ALIGNMENTS

RESULT	1
Q9KFV3	
ID	Q9KFV3 PRELIMINARY; PRT; 177 AA.
AC	Q9KFV3;
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
GN	BH0368.
OS	Bacillus halodurans.
OC	Bacteria; Firmicutes; Bacillus/clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_TaxID=86665;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-C-125 / JCM 9153;
RX	MEDLINE=20512582; PubMed=11058132;
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA	Horikoshi K.;
RT	*Complete genome sequence of the alkaliphilic bacterium Bacillus
RT	halodurans and genomic sequence comparison with Bacillus subtilis.;
RL	Nucleic Acids Res. 28:4317-4331(2000).
DR	EMBL; AP001508; BAB04087.1; -
DR	InterPro; IPR003232; dCTP_deaminase.
DR	InterPro; IPR001428; dUTPase.
DR	Pfam; PF00692; dUTPase; 1
DR	ProDom; PD004900; dCTP_deaminase; 1.
KW	Complete proteome.
SQ	SEQUENCE 177 AA; 19897 MW; D15AE75387847E2B CRC64;

Query Match	30.1%;	Score 242.5;	DB 2;	Length 177;
Best Local Similarity	37.2%;	Pred. No. 4.1e-14;		
Matches	58;	Conservative	30;	Mismatches 47; Indels 21; Gaps 6;
Qy	10	KEILIEPFSEESLOPAGYDLRVGREAFV-----KGKLDIVKEK-----EGKVYIPP	54	
Db	15	KELITITLPEEQIQPASVDLRGLGPH-FVTIDDSKEAIVSEFPIRYREWTSTDEIVLPP	73	
Qy	55	REYALITLRIKLPDDVMGDMGKTRSSLAREGV-IGSFANVDPGWDGNLTLMLYNASNEP	113	

```

Db      74 HTPELLATTWETVKPLNHLTAFTVEGRSSVGRIGLFIQTNGAGWDPGFNGQITILELFANRLP   133
QY      114 VELRYGSRVFQIAFRLEGPARNPYRGNY---QGST 146
           :||| | | : | : ||| | | : |||
Db      134 IELPIGRRICQLVFAEVTGEVA-PYGKYLFOKGAT 168

RESULT    2
071028
ID       O71028          PRELIMINARY;              PRT;        158 AA.
AC       071028;
DT       01-AUG-1998 (TREMBRel. 07, Created)
DD       01-AUG-1998 (TREMBRel. 07, Last sequence update)
DE       01-JUN-2001 (TREMBRel. 17, Last annotation update)
DE       DUTPASE (EC 3.6.1.23) (DUTP PYROPHOSPHATASE) (DEOXYURIDINE-
OS       TRIPHOSPHATASE).
DS       Archaeal virus SIRV.
OX       Viruses; unclassified viruses.
NCBI_Taxid=66287;
RN       [1]
RP       SEQUENCE FROM N.A.
RX       MEDLINE=98165772; Pubmed=9497317;
RA       Prangishvili D., Klenk H.P., Jakobs G., Schmielechen A., Hanselmann C.,
RA       Holz L.I., Zillig W.;
RT       "Biochemical and phylogenetic characterization of the dUTPase from the
RT       archaeal virus SIRV.";
RL       J. Biol. Chem. 273:6024-6029(1998).
CC       -I- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
CD       EMBL; AF022221; AAC15873.1; -.
DR       InterPro: IPR001428; dUTPase.
DR       Pfam: IPR003232; dCTP_deaminse.
DR       ProDom: PD00692; dUTPase; 1.
DR       ProDom: PD004900; dCTP_deaminse; 1.
KW       Hydrolyase.
SQ       SEQUENCE     158 AA; 17906 MW;  5BA47765E0889190 CRC64;
```

[illegible]

RESULT	3	
Q9CB17		
ID	Q9CB17	
AC	PRELIMINARY;	PRT: 190 AA.
DC	Q9CB17;	
DT	01-JUN-2001 (TREMBrel. 17, Created)	
DT	01-JUN-2001 (TREMBrel. 17, Last sequence update)	
DT	01-JUN-2001 (TREMBrel. 17, Last annotation update)	
DT	01-JUN-2001 (TREMBrel. 17, Last annotation update)	
DE	PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.	
GN	ML2507.	
GO	Mycobacterium leprae.	
OS	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OC	NCBI_TaxID=11769;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RP	STRAIN-TN;	
RC	MEDLINE=21128732; PubMed=11234002;	
RC	Cole S.T., Eaglemeier K., Parkhill J., James K.D., Thomson N.R.,	
RA	Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,	

	Mungall K., Basham D., Brown D., Chillingworth T., Connor R., Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N., Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S., Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M., Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R., Barrett B.G.;
RA	"Massive gene decay in the leprosy bacillus.";
RL	Nature 409:1007-1011(2001).
RT	EMBL; AL583925; CAC32024.1; -.
DR	InterPro: IPR001232; dCTP_deaminase.
DR	InterPro: IPR001428; dUTPase.
DR	Pfam: PF00692; dUTPase; 1.
DR	ProDom: PD004900; dCTP_deaminase; 1.
KW	Complete proteome.
SQ	SEQUENCE 190 AA; 20820 MW; 1CA936700500B6FA CRC64;
	Query Match 24.5%; Score 197.5; DB 2; Length 190; Best Local Similarity 30.6%; Prod No 4.se-10; Matches 53; Conservative 32; Mismatches 59; Indels 29; Gaps
Qy	1 MLPPDWKIREIL-----IPFFSEESLOPAGYDLRV-----GSAFVK 38 : : : : : : : : : :
Dd	1 MLLSDRDLRAEITAGRFSIDFDTLVQPSSIDVLDCMFRVFNTNRYTHIDPARQQDEL 60 : : : : : : : : : :
Qy	39 GKLIIDKEKGKVTPPREYALIIPLRIKLDDVMGMKRISSLARGVI--GSFAWDVP 96 : : : : : : : : : :
Dd	61 TSLSVELVDGGEPFLVHPGGFVLGSTLFTLPEDLAGLEGKSLSGLRGILLTHTSHAGFDP 120 : : : : : : : : : :
Qy	97 GWGDGNLTLMYNASNEPVELRYGERFVOIAIFIRLEGPARNPY-----RGNYQG 144 : : : : : : : : : :
Dd	121 GFCHGHITLESNVANLRPTLWPNGMKIGQLCVLRUTSPAEPHYGSASAKSYQG 173 : : : : : : : : : :
RESULT	4
ID	Q9PN07 PRELIMINARY; PRT; 186 AA.
O	AQN007;
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	POSSIBLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13).
DN	CDCl OR CuI292.
OS	Campylobacter jejuni.
OC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OX	Campylobacter.
QC	NCBI_TaxID=197;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NCTC 11168.
RX	MEDLINE=J0150912; PubMed=10688204;
RA	Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C., Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S., Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W., Quail M.A., Rajandream M.A., Rutherford K.M., Van Vleet A.H.M., Whitehead S., Barrett B.G.;
RA	"The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences."
RL	Nature 403:665-668(2000).
RR	EMBL; AL139078; CAB73719.1; -.
DR	InterPro: IPR003232; dCTP_deaminase.
DR	ProDom: PD004900; dCTP_deaminase; 1.
KW	Complete proteome.
SQ	SEQUENCE 186 AA; 20699 MW; 68DC660FB28FD33C CRC64;

```

Query Match      23.5%; Score 189.5; DB 2; Length 186;
Best Local Similarity 29.4%; pred. No. 2.3e-09;
Matches 53; Conservative 34; Mismatches 58; Indels 35; Gaps 7;

QY 5 DWKIRKEIL---IEPFSEES-----LQPGAYDLRVG-----AF 36
      : ||| | : ||| | : ||| | ||| |

```

Db 7 NW-IRKMALEHKMTEPCFENAGIKGVVSYGLSSYGYDIRVGRFEKFTNVNSTVDPKNE 65
 QY 37 VKGLIDVEKGV-VIPPREYALILTLERIKLPDDYMGDMKIRSSLAREGVIGSFAWD 95
 Db 66 VEENVDF--EGDVCIIPANSFALARTIEYFMPDNVLALCLGKSTYARCGIIVNTPFE 123
 QY 96 PGWGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY---RGNYGSTRLAFSK 152
 Db 124 PGFEGHITIEISNTPLPAKIYANEGIAQVLFLOGDEKCDTYYDKKGYQAQTGITLPR 183

RESULT 5
 Q9HSG3 ID Q9HSG3 PRELIMINARY; PRT; 195 AA.
 AC Q9HSG3
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
 GN DTD OR VNG0245G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.W., Liang P., Riley M., Hood L., Dassarma S.;
 RA "Genome sequence of Halobacterium species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AE004988; AAG18843.1; -
 DR InterPro; IPR003232; dCTP_deaminase.
 DR Pfam; PF00692; dUTPase; 1.
 DR ProDom; PD004900; dCTP_deaminase; 1.
 KW Complete proteome.
 SQ SEQUENCE 195 AA; 21542 MW; 86CCDA4D48BC0C9D CRC64;

QY 1 MLPLDWMKIRK-----EILIPFSESL--OPAGYDLRVGRE--AFVKGKL----- 41
 Db 1 MILSDQILARLADGLAIEPLEVDVLOVQSPASVDVLRGRFLFEFRANVPICHPNREDE 60

QY 42 ID-----VEKEGVVIPPREYALILTLERIKLPDDYMGDMKIRSSLAREGVI--GSFA 92
 Db 61 VDEVVTVETVWDGDEFILHPGDFVLGTTKERVEPRDLVAQVCEGRSSILGLAVVHHATAG 120

QY 93 WVDPGWGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPYRGNQGS 145
 Db 121 FIDPGFNGRVTLESLNGLKVPVVALTPEMRISQLVFTLTSPADRPY-GDERGS 172

RESULT 6
 Q9HYC9 ID Q9HYC9 PRELIMINARY; PRT; 188 AA.
 AC Q9HYC9
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
 GN PA3480.
 OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004769; AAG06868.1; -
 DR InterPro; IPR003232; dCTP_deaminase.
 DR InterPro; IPR001428; dUTPase.
 DR Pfam; PF00692; dUTPase; 1.
 DR ProDom; PD004900; dCTP_deaminase; 1.
 KW Complete proteome.
 SQ SEQUENCE 188 AA; 21154 MW; 9A811B5F17B2A996 CRC64;

Query Match 21.3%; Score 172; DB 2; Length 188;
 Best Local Similarity 26.3%; Pred. No. 8.2e-08;
 Matches 45; Conservative 32; Mismatches 62; Indels 32; Gaps 4;

QY 13 LIEPFSESLQA-----GYDLRVGRE-----AFVKGKLIDV 44
 Db 18 MIEPFVERQVGRGDSRVSISGVSSGYDVRCAAEFVFTNIHSAVDPKPFDEKSFVDI 77

QY 45 EKEGVVIPPREYALILTLERIKLPDDYMGDMKIRSSLAREGVIGSFAWDPGWGNLT 104
 Db 78 NSD-VCIIPNSFALARTVEYFPRDVLITCLGKSTYARCGIIVNTPLEPEWGHVTL 136

QY 105 MLYNASNEPVELRYGERFVQIAFIRLEGPARNPYR--GNYQGSTRLAFSK 152
 Db 137 EFSNTNLPKAIYANEGVAQMLFQSDACEVSKDRGKYGQGRGVTLPK 187

RESULT 7
 Q9PFB6 ID Q9PFB6 PRELIMINARY; PRT; 191 AA.
 AC Q9PFB6
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
 GN XF0762.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OX Xylella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C.
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.F., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meldonis J., Setubal J.C.,
 RA "the genome sequence of the plant pathogen Xylella fastidiosa";
 BL Nature 406:151-159(2000).
 DR EMBL: AE003917; AAF83572.1; --
 DR InterPro: IPR003232; dCTP_deaminase; 1.
 DR Problem: PD004900; dCTP_deaminase; 1.
 KW Complete proteome.
 SQ SEQUENCE 191 AA; 21531 MW; D7B23653F94B3649 CRC64;

Query Match 20.4%; Score 164.5; DB 2; Length 191;
 Best Local Similarity 30.5%; Pred. No. 3.9e-07;
 Matches 43; Conservative 25; Mismatches 52; Indels 21; Gaps 4;
 QY 26 GYDLRVGR-----AFVKGKLDVKEGKVVIPPREVALIILTLERIKLPD 70
 DB 45 GYDVRCSEFKFTINSTVDPKQDFNGSFIDVESD-VCIIPNPSFALARTIEYFRIPR 103
 QY 71 DVMGDMKIRSLAREGVGSPAWDPGWDGNLTMLYNASNEPVELRYGERFVOIAFIRL 130
 DB 104 NVLVICLGKSTYARCGIIVNVNTPLEPEWEGHVTLEFSNTTLPARIYANEGVAQMLFQA 163
 QY 131 --EGPARNPYR--GNYOGST 146
 DB 164 DPDDVCTSYDRNGKVGQQT 184

RESULT 8
 QYIG32 PRELIMINARY; PRT; 163 AA.
 AC QYIG32
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE 163AA LONG HYPOTHETICAL DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
 GN APE0069.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funabashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RA "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1";
 RL DNA Res. 6:83-101(1999)
 DR EMBL: AP000058; BAA78978.1; --
 DR InterPro: IPR003232; dCTP_deaminase.
 DR Problem: PD004900; dCTP_deaminase; 1.
 KW Complete proteome.
 SQ SEQUENCE 163 AA; 17384 MW; 4AA22FB0D8802F49 CRC64;

Query Match 19.7%; Score 159; DB 1; Length 163;
 Best Local Similarity 32.7%; Pred. No. 9.8e-07;
 Matches 48; Conservative 19; Mismatches 62; Indels 18; Gaps 5;
 QY 13 LIEFPSESLOPAGYDLRVGR-----EAFVKGKLDVKEGKVVIPPREYA 58
 DB 17 VVKHNGAIOACGVDSLVSGBIESLADAGFLGEDKIMPKGDRIQCE-YGVCELEPGAYR 75
 QY 59 LILTLERIKPDDVGMKIRSLAREG-VIGSPAWDPGWDGNLTMLYNASNEPVELR 117
 DB 76 LRFN-EVVSIPPGHVGCFPRSSLLRMGCYLCGCAVW-DPGYTGRCQAMLLVANPGLRLE 133
 QY 118 YGERFVOIAFIRLEGPARNPYRQYQ 144
 DB 134 MGSRIQLVVARVEGPLTSYKGYQ 160
 RESULT 9
 QYJRE8 PRELIMINARY; PRT; 188 AA.
 AC QYJRE8
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL PROTEIN NMA1060 (DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE,
 DE PUTATIVE).
 GN NMA1060 OR NMB0849.
 OS Neisseria meningitidis (serogroup A), and
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699, 491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Doherty K., Feilwell T., Hamlin N., Holtroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 RL meningitidis Z2491";
 RL Nature 404:502-506(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 RC MC58";
 RL Science 287:1809-1815(2000).
 DR EMBL: AL162755; CAB84324.1; --
 DR EMBL: AE002438; AAF41260.1; --
 DR TIGR: NMB0849; --
 DR InterPro: IPR003232; dCTP_deaminase.
 DR Problem: PD004900; dCTP_deaminase; 1.
 DR InterPro: IPR001428; dUTPase.
 DR Fram: PF00652; dUTPase; 1.
 DR Problem: PD004900; dCTP_deaminase; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 188 AA; 21295 MW; 2CA4459F4E78FDF6 CRC64;

Query Match 19.2%; Score 155; DB 2; Length 188;
 Best Local Similarity 24.1%; Pred. No. 2.7e-06;
 Matches 41; Conservative 32; Mismatches 67; Indels 30; Gaps 4;

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE RETROPEPSIN FAMILY.

CC EMBL: AF247394; AAC02702.1; -
CC InterPro: IPR001969; Asp_protease.
CC InterPro: IPR001995; Asp_protr_retrov.
CC InterPro: IPR001428; dUTPase.
CC InterPro: IPR002156; RNaseH.
CC InterPro: IPR001584; Rve.
CC InterPro: IPR000477; RVTse.
CC Pfam: PF006692; dUTPase; 1.
CC Pfam: PF00075; rnaaseH; 1.
CC Pfam: PF00665; rve; 1.
CC Pfam: PF00077; rvp; 1.
CC Pfam: PF00078; rvt; 1.
CC ProDom: PD000946; dUTPase; 1.
CC PROSITE: PS00141; ASP_PROTEASE; 1.
CC PROSITE: PS01175; ASP_PROT_RETROV; 1.
CC Aspartyl protease; Hydrolase; RNA-directed DNA polymerase.
FT NON_TER 1
SQ SEQUENCE 1146 AA; 129496 MW; 79B80920D8D20AB5 CRC64;

Query Match 15.7%; Score 126.5; DB 12; Length 1146;
Best Local Similarity 23.0%; Pred. No. 0.0097;
Matches 32; Conservative 38; Mismatches 50; Indels 19; Gaps 2;

QY 7 KIRKEILI---EFSESLOPAGYDLRVGEAFVKGKLDIVKEGKVIPPREYALILT 62
DB 736 KIKEIMLAYGTOGKERDEDAFDLCVPYD-----IMIPVSTKIPT 780
QY 63 LERKLPDDVGMKIRSSLAKEGVIGSFVWPGDGNLTMLYNASNEPVELRYGERF 122
DB 781 DVKLRVPNSFGVGTGKSSMAKQGLLGIIDEGYTGTEIQTICNTKSNIKLIEGKF 840
QY 123 VQAFIRLEGPARPYRGN 141
DB 841 AOLILQHHSNSRQPDEN 859

RESULT 13
O29157 ID 029157 PRELIMINARY; PRT; 168 AA.
AC O29157 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE, PUTATIVE.
GN AF1108.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
SQ SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Katchum K.A., Dodson R.J., Gunn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kurlavsky A.R., Graham D.E., Kyriakides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL: AE001027; AAB90130.1; -
DR TIGR: AF1108; -
DR InterPro: IPR003232; dCTP_deaminase.

DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hypothetical protein; Complete Proteome.
SQ SEQUENCE 168 AA; 19227 MW; D99EE0FB943869D6 CRC64;

Query Match 15.6%; Score 126; DB 1; Length 168;
Best Local Similarity 29.4%; Pred. No. 0.00089;
Matches 45; Conservative 25; Mismatches 65; Indels 18; Gaps 5;

QY 8 IRKEILIEPF--SEESLOPAGYDLRV-----GREAFVKGK-----LIDVEKEGKV 51
DB 13 IQKEGLIRDYVDLETOIQPNFGDCTLSRVYLRGCRVDFDNRRLPELEVEFEFDWVY 72
QY 52 IPPREYALILTLEKILPDDVGMKIRSSLAKEGVIGSFVWPGDGNLTMLYNASN 111
DB 73 LPKGVYRAKLN-EVVRGNDIMAIAPRSTLIRCGANVLTAVWDAGYEGRSEVSIHVND 131
QY 112 EPVELRYGERFVQAFIRLEGPARPYRGN 144
DB 132 YGIWLSRNARIQLVFIRLSSPTKG-YEGVYKG 163

RESULT 14
O9WHF3 ID O9WHF3 PRELIMINARY; PRT; 170 AA.
AC O9WHF3 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 18.5 KDA PROTEIN.
OS Agrotis segetum granulosis virus (AsGV) (Agrotis segetum
OS granulovirus).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX NCBI_TaxID=10464;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=XJASGV;
RA Xiulian A., Ning W., Wei Z., Yuhu S.;
RT "The sequence analysis of two BamHI fragments of Agrotis segetum
RT granulovirus DNA.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF130846; AAD34382.1; -
DR HSPB; P06968; IEUW.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD000946; dUTPase; 1.
KW Hypothetical protein.
SQ SEQUENCE 170 AA; 18473 MW; 812982A6E508ECOD CRC64;

Query Match 15.6%; Score 125.5; DB 12; Length 170;
Best Local Similarity 27.5%; Pred. No. 0.00099;
Matches 39; Conservative 31; Mismatches 57; Indels 15; Gaps 4;

QY 29 LRVGREAFVKGK----IDVEKEGKVIPPREYALILTLEKILPDDVGMKIRSSLA- 83
DB 21 LRKDAHAFVRSVGAAGLDLASVENIQIPGENALVPTGLSIEIPNDICYGRIAPRGLAL 80
QY 84 REGVIGSFVWPGDGNLTMLYNASNPEVELRYGERFVQAFIRLEGP-----A 134
DB 81 RYSVSVHAGVIDPDYRGHLQVLEFNHGKTKMEILKGRIOALQVCEKIRFPEAVEKPKLSE 140
QY 135 RNPYRGNYGQSTRLAFSKRKL 156
DB 141 TDRAEGGF-GSTGIASQQQEEI 161

RESULT 15
O9W7U6 ID O9W7U6 PRELIMINARY; PRT; 1138 AA.
AC O9W7U6;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:19 ; Search time 140.03 Seconds
(without alignments)
8.993 Million cell updates/sec

Title: US-08-957-709-11

Perfect score: 17
Sequence: 1 GAILLPDMKIRKEILLIE 17

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_ll01.*

- 1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	82.4	15	AAW72845	Polymerase enhanci
2	14	82.4	156	AAW72847	Polymerase enhanci
3	8	47.1	18	AAW72846	Polymerase enhanci
4	7	41.2	15	AAW72870	Polymerase enhanci
5	6	41.2	782	AAB93202	Human protein sequ
6	6	35.3	93	AAG56384	Arabidopsis thalia
7	6	35.3	93	AAG61374	Arabidopsis thalia
8	6	35.3	100	AA556383	Arabidopsis thalia
9	6	35.3	100	AAG61373	Arabidopsis thalia
10	6	35.3	110	AA556382	Arabidopsis thalia
11	6	35.3	110	AAG61372	Arabidopsis thalia

12	6	35.3	151	21	AA41187	Human ORFX ORF951
13	6	35.3	154	21	AAG16696	Arabidopsis thalia
14	6	35.3	182	22	AAB93696	Human protein sequ
15	6	35.3	200	21	AAB25223	Eucalyptus grandis
16	6	35.3	225	21	AAG20301	Arabidopsis thalia
17	6	35.3	253	20	AA33724	Photobacterium
18	6	35.3	256	21	AAG16695	Arabidopsis thalia
19	6	35.3	269	21	AAG20300	Arabidopsis thalia
20	6	35.3	275	21	AAG16694	Arabidopsis thalia
21	6	35.3	303	21	AAG40736	Zea mays protein f
22	6	35.3	303	21	AAG43821	Arabidopsis thalia
23	6	35.3	303	21	AAG44221	Arabidopsis thalia
24	6	35.3	311	19	AAW79286	Human phosphatidic
25	6	35.3	339	22	AAG72102	Human olfactory re
26	6	35.3	349	22	AAB46747	R. marinus bacteri
27	6	35.3	355	21	AAG20299	Arabidopsis thalia
28	6	35.3	360	21	AAG52354	Arabidopsis thalia
29	6	35.3	401	22	AA42041	Human polypeptide
30	6	35.3	407	21	AAG40735	Zea mays protein f
31	6	35.3	438	20	AA36904	Protein involved i
32	6	35.3	500	22	AAB61315	Human transcriptio
33	6	35.3	511	18	AAW34560	Thermococcus chito
34	6	35.3	511	19	AAW49864	Thermococcus GC74-
35	6	35.3	570	17	AA96300	Foki insertion mut
36	6	35.3	573	17	AA96299	Foki insertion mut
37	6	35.3	577	15	AA58759	Foki amino acid se
38	6	35.3	579	16	AA72810	Foki. Flavobacter
39	6	35.3	581	17	AA88594	Foki insertion mut
40	6	35.3	581	17	AA88593	Foki insertion mut
41	6	35.3	583	11	AA806978	Modified enzyme fr
42	6	35.3	584	20	AA37374	Protein which is s
43	6	35.3	585	17	AA88595	Foki insertion mut
44	6	35.3	589	17	AA88596	Foki insertion mut
45	6	35.3	595	17	AA88599	Foki insertion mut

ALIGNMENTS

RESULT 1
AAW72845
ID AAW72845 standard; Peptide; 15 AA.
XX
AC AAW72845;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 component N-terminal peptide.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /label= Gly, Ala, Ile
XX
PN W09842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX
PA (STRA-) STRATAGENE.
XX
PI Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
XX

PT Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 XX use in amplification, sequencing and replication
 PS Claim 17; Page 33; 161pp; English.
 XX This peptide comprises an N-terminal peptide of the p45 component
 CC of the polymerase enhancing factor (PEF) of *Pyrococcus furiosus*
 CC DSM 3638. It has been obtained from isolated p45; a full-length
 CC sequence is provided in AAW72847. PEF, the predominant components
 CC of which are p45 and p50 (see AAW72844) proteins, enhances the
 CC activity of *P. furiosus* DNA polymerase, thereby providing
 CC replication products of greater length and purity. p45 is a
 CC dUTPase and possesses polymerase enhancing activity. The invention
 CC provides novel extracts, proteins and complexes that improve the
 CC polymerisation activity of nucleic acid polymerases. These
 CC complexes may include proteins including the p45 N-terminal
 CC peptide. Also included are methods for identifying compositions
 CC with polymerase enhancing activity, for purifying and using these
 CC compositions, and specific extracts, proteins and complexes that
 CC function to enhance polymerase activity. Nucleic acid polymerase
 CC reactions can be enhanced (claimed) by mixing a nucleic acid
 CC template, at least 1 polymerase and a composition having polymerase
 CC enhancing activity. Kits are provided for replicating nucleic
 CC acids. The kits can be used in site-directed mutagenesis, nucleic
 CC acid sequencing or amplification (preferably PCR or RT-PCR).
 XX Sequence 15 AA;

Query Match 82.4%; Score 14; DB 19; Length 15;
 Best Local Similarity 100.0%; Pred. No. 5e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 4 LLPDKIRKEILIE 17
 Db 2 llpdwkkirkellie 15
 RESULT 2
 AAW72847
 ID AAW72847 standard; Protein: 156 AA.
 XX AAW72847;
 AC AAW72847;
 XX 01-MAR-1999 (first entry)
 DT Polymerase enhancing factor p45 (dUTPase) component.
 DE Polymerase enhancing factor: PEF; dUTPase; DNA polymerase; PCR;
 XX amplification; sequencing; replication.
 KW *Pyrococcus furiosus* strain DSM 3638.
 OS
 XX Key Location/Qualifiers
 XX Peptide 2..15
 FT /note= "N-terminal peptide used to generate
 FT primers"
 XX WO9842860-A1.
 PN 01-OCT-1998.
 PD 20-MAR-1998; 98WO-US05497.
 XX 24-OCT-1997; 97US-0957709.
 PR 21-MAR-1997; 97US-0822774.
 XX (STRA-) STRATAGENE.
 PA Hansen CJ, Hogrefe H;
 XX WPI; 1998-542284/46.
 XX

DR N-PSDB; AAV63860.
 XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 XX use in amplification, sequencing and replication
 PS Claim 17; Page 43; 161pp; English.
 XX This is the amino acid sequence of the p45 component of the
 CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.
 CC The sequence is predicted from a DNA sequence (see AAV63860) obtained
 CC from genomic DNA by PCR. p45 and p50 (see AAW72844) are the
 CC predominant components of PEF, which acts to enhance the activity of
 CC *P. furiosus* DNA polymerase. p45 functions as a dUTPase, and can be
 CC used to enhance nucleic acid replication, polymerisation or PCR
 CC reactions. The invention provides novel extracts, proteins and
 CC complexes that improve the polymerisation activity of nucleic acid
 CC polymerases, as well as DNA constructs and antibodies. Also
 CC included are methods for identifying compositions with PEF activity,
 CC for purifying and using these compositions, and specific extracts,
 CC proteins and complexes that function to enhance polymerase activity.
 CC Nucleic acid polymerase reactions can be enhanced (claimed) by
 CC mixing a nucleic acid template, at least 1 polymerase and a
 CC composition having polymerase enhancing activity. Kits are
 CC provided for replicating nucleic acids, for site-directed
 CC mutagenesis, for nucleic acid sequencing or amplification
 CC (preferably PCR or RT-PCR).
 XX Sequence 156 AA;

Query Match 82.4%; Score 14; DB 19; Length 156;
 Best Local Similarity 100.0%; Pred. No. 4.1e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 4 LLPDKIRKEILIE 17
 Db 2 llpdwkkirkellie 15
 RESULT 3
 AAW72846
 ID AAW72846 standard; Peptide: 8 AA.
 XX AAW72846;
 AC AAW72846;
 XX 01-MAR-1999 (first entry)
 DT Polymerase enhancing factor p45 component N-terminal peptide.
 DE Polymerase enhancing factor: PEF; dUTPase; DNA polymerase; PCR;
 XX amplification; sequencing; replication.
 KW *Pyrococcus furiosus* strain DSM 3638.
 OS
 XX WO9842860-A1.
 PN 01-OCT-1998.
 PD 20-MAR-1998; 98WO-US05497.
 XX 24-OCT-1997; 97US-0957709.
 PR 21-MAR-1997; 97US-0822774.
 XX (STRA-) STRATAGENE.
 PA Hansen CJ, Hogrefe H;
 XX WPI; 1998-542284/46.
 XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 XX use in amplification, sequencing and replication

XX PS Claim 17; Page 33; 161pp; English.

XX CC This peptide comprises an N-terminal peptide of the P45 component

XX CC of the polymerase enhancing factor (PEF) of *Pyrococcus furiosus*

XX CC DSM 3638. It has been obtained from isolated P45. PCR primers

XX CC (see AAV63861-64) based on the peptide were used to amplify P45 DNA

XX CC (see AAV63860). A full-length P45 sequence is provided in AAW72847.

XX CC PEF, a predominant component of which is P45 protein, enhances the

XX CC activity of *P. furiosus* DNA polymerase, thereby providing

XX CC replication products of greater length and purity. P45 is a

XX CC dUTPase and possesses polymerase enhancing activity. The invention

XX CC provides novel extracts, proteins and complexes that improve the

XX CC polymerisation activity of nucleic acid polymerases. These

XX CC complexes may include proteins including the P45 N-terminal

XX CC peptide. Also included are methods for identifying compositions

XX CC with polymerase enhancing activity, for purifying and using these

XX CC compositions, and specific extracts, proteins and complexes that

XX CC function to enhance polymerase activity. Nucleic acid polymerase

XX CC reactions can be enhanced (claimed) by mixing a nucleic acid

XX CC template, at least 1 polymerase and a composition having polymerase

XX CC enhancing activity. Kits are provided for replicating nucleic

XX CC acids. The kits can be used in site-directed mutagenesis, nucleic

XX CC acid sequencing or amplification (preferably PCR or RT-PCR).

XX SQ Sequence 8 AA;

Query Match 47.1%; Score 8; DB 19; Length 8;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PDWKIRKE 13
 |||||
 Db 1 pdwkirke 8

RESULT 4
 AAW72870
 ID AAW72870 standard; Peptide; 15 AA.
 XX AC AAW72870;
 XX XX 01-MAR-1999 (first entry)
 DT DT Polymerase enhancing factor P45 protein N-terminal peptide.
 DE DE
 XX KW Polymerase enhancing factor; PEF; DNA polymerase; dUTPase; PCR;
 KW amplification; sequencing; replication.
 XX OS *Pyrococcus furiosus* strain DSM 3638.
 XX XX
 XX FH Key Location/Qualifiers
 FT Misc-difference 2 /label= Leu, Tyr
 FT Misc-difference 3 /label= Leu, Val
 FT Misc-difference 4 /label= Arg, Pro
 FT FT
 XX PN W09842860-A1.
 XX PD
 XX XX 01-OCT-1998.
 XX XX 20-MAR-1998; 98WO-US05497.
 XX XX 24-OCT-1997; 97US-0957709.
 XX XX 21-MAR-1997; 97US-0822774.
 XX XX (STRA-) STRATAGENE.
 XX XX Hansen CJ, Hogrefe H;
 PI
 XX

DR WPI; 1998-542284/46.

XX PT Polymerase enhancing factor proteins, extracts and complexes -

PT improve the polymerisation activity of nucleic acid polymerases, for

XX use in amplification, sequencing and replication

XX PS Example 5; Page 33; 161pp; English.

XX CC This is an N-terminal peptide of the P45 component of the

XX CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.

XX CC PEF, the predominant components of which are P50 (see AAW72844) and

XX CC P45 (see AAW72847), enhances the activity of *P. furiosus* DNA

XX CC polymerase. P45 functions as a dUTPase. The invention provides

XX CC novel extracts, proteins and complexes that improve the

XX CC polymerisation activity of nucleic acid polymerases. These can be

XX CC used to improve nucleic acid replication, polymerisation and

XX CC amplification (especially in PCR or RT-PCR).

XX SQ Sequence 15 AA;

Query Match 41.2%; Score 7; DB 19; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 RKEILIE 17
 |||||
 Db 9 rkeilie 15

RESULT 5
 AAB93202
 ID AAB93202 standard; Protein; 782 AA.
 XX AC AAB93202;
 XX XX 26-JUN-2001 (first entry)
 DT DT Human protein sequence SEQ ID NO:12168.
 DE DE
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX XX 28-JUL-2000; 2000EP-0116126.
 XX XX 29-JUL-1999; 99JP-0248036.
 XX XX 27-AUG-1999; 99JP-0300253.
 XX XX 11-JAN-2000; 2000JP-0118776.
 XX XX 02-MAY-2000; 2000JP-0183767.
 XX XX 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX XX
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR WPI; 2001-318749/34.
 XX XX
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

XX PT full-length cDNAs -

XX PS Claim 8; SEQ ID 12168; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesising 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

SQ Sequence 782 AA;

Query Match 41.2%; Score 7; DB 22; Length 782;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KIRKEIL 15
 |||||
 Db 51 kirkeil 57

RESULT 6

AAAG56384
 ID AAG56384 standard; Protein; 93 AA.

AC AAG56384;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 72469.

XX protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137252.
 PR 03-JUN-1999; 99US-0137558.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140981.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145066.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145087.

```
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 6; DB 21; Length 93;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ILLPDW 8
Db 28 illpdw 33

RESULT 7
AAG61374
ID AAG61374 standard; Protein; 93 AA.
XX AAG61374;
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 79599.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
```

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134708.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158372.
PR 12-OCT-1999; 99US-0158569.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 35, 38; Score 6; DB 21; Length 93;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ILLPDW 8
|||||
Db 28 illpdw 33

RESULT 8
AAG56383
ID AAG56383 standard; Protein; 100 AA.

XX AC AAG56383;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 72468.

DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 6; DB 21; Length 100;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 3 ILLPDW 8
 Db 35 illpdw 40
 |||||

RESULT 9
 AAG61373
 ID AAG61373 standard; Protein; 100 AA.
 XX
 AC AAG61373;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 79598.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 OS Arabidopsis thaliana.
 XX
 DN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0128845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 05-MAY-1999; 99US-0132486.
 PR 05-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 01-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 6; DB 21; Length 100;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 ILLPDW 8

```
Db          |||||
35  illpdw 40

RESULT 10
ID  AAG56382 standard; Protein; 110 AA.
XX
AC  AAG56382;
XX
DT  18-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 72467.
XX
KW  protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-0301439.
XX
PR  25-FEB-1999; 99US-0121825.
PR  05-MAR-1999; 99US-0123180.
PR  09-MAR-1999; 99US-0123548.
PR  23-MAR-1999; 99US-0125788.
PR  25-MAR-1999; 99US-0126264.
PR  29-MAR-1999; 99US-0126785.
PR  01-APR-1999; 99US-0127462.
PR  06-APR-1999; 99US-0128234.
PR  08-APR-1999; 99US-0128714.
PR  16-APR-1999; 99US-0129845.
PR  19-APR-1999; 99US-0130077.
PR  21-APR-1999; 99US-0130510.
PR  23-APR-1999; 99US-0130891.
PR  28-APR-1999; 99US-0131449.
PR  30-APR-1999; 99US-0132048.
PR  04-MAY-1999; 99US-0132407.
PR  05-MAY-1999; 99US-0132484.
PR  06-MAY-1999; 99US-0132485.
PR  07-MAY-1999; 99US-0132486.
PR  11-MAY-1999; 99US-0132863.
PR  14-MAY-1999; 99US-0134256.
PR  14-MAY-1999; 99US-0134218.
PR  14-MAY-1999; 99US-0134219.
PR  14-MAY-1999; 99US-0134221.
PR  18-MAY-1999; 99US-0134370.
PR  19-MAY-1999; 99US-0134768.
PR  20-MAY-1999; 99US-0134941.
PR  21-MAY-1999; 99US-0135124.
PR  24-MAY-1999; 99US-0135353.
PR  25-MAY-1999; 99US-0135629.
PR  27-MAY-1999; 99US-0136021.
PR  28-MAY-1999; 99US-0136782.
PR  01-JUN-1999; 99US-0137222.
PR  03-JUN-1999; 99US-0137528.
PR  04-JUN-1999; 99US-0137502.
PR  07-JUN-1999; 99US-0137724.
PR  08-JUN-1999; 99US-0138094.
PR  10-JUN-1999; 99US-0138540.
PR  10-JUN-1999; 99US-0138847.
PR  14-JUN-1999; 99US-0139119.
PR  16-JUN-1999; 99US-0139452.
PR  16-JUN-1999; 99US-0139453.
PR  17-JUN-1999; 99US-0139492.
PR  18-JUN-1999; 99US-0139454.
PR  18-JUN-1999; 99US-0139455.
PR  18-JUN-1999; 99US-0139456.
PR  18-JUN-1999; 99US-0139457.
PR  18-JUN-1999; 99US-0139458.
PR  18-JUN-1999; 99US-0139459.
PR  18-JUN-1999; 99US-0139460.
PR  18-JUN-1999; 99US-0139461.
PR  18-JUN-1999; 99US-0139462.
PR  18-JUN-1999; 99US-0139463.
PR  18-JUN-1999; 99US-0139750.
PR  18-JUN-1999; 99US-0139763.
PR  21-JUN-1999; 99US-0139817.
PR  22-JUN-1999; 99US-0139899.
PR  23-JUN-1999; 99US-0140353.
PR  23-JUN-1999; 99US-0140354.
PR  24-JUN-1999; 99US-0140695.
PR  28-JUN-1999; 99US-0140823.
PR  29-JUN-1999; 99US-0140991.
PR  30-JUN-1999; 99US-0141287.
PR  01-JUL-1999; 99US-0141842.
PR  01-JUL-1999; 99US-0142154.
PR  02-JUL-1999; 99US-0142055.
PR  06-JUL-1999; 99US-0142390.
PR  08-JUL-1999; 99US-0142803.
PR  09-JUL-1999; 99US-0142920.
PR  12-JUL-1999; 99US-0142977.
PR  13-JUL-1999; 99US-0143542.
PR  14-JUL-1999; 99US-0143624.
PR  15-JUL-1999; 99US-0144005.
PR  16-JUL-1999; 99US-0144085.
PR  16-JUL-1999; 99US-0144086.
PR  19-JUL-1999; 99US-0144325.
PR  19-JUL-1999; 99US-0144331.
PR  19-JUL-1999; 99US-0144332.
PR  19-JUL-1999; 99US-0144333.
PR  19-JUL-1999; 99US-0144334.
PR  19-JUL-1999; 99US-0144335.
PR  20-JUL-1999; 99US-0144352.
PR  20-JUL-1999; 99US-0144632.
PR  20-JUL-1999; 99US-0144884.
PR  21-JUL-1999; 99US-0144814.
PR  21-JUL-1999; 99US-0145086.
PR  21-JUL-1999; 99US-0145088.
PR  22-JUL-1999; 99US-0145085.
PR  22-JUL-1999; 99US-0145087.
PR  22-JUL-1999; 99US-0145089.
PR  22-JUL-1999; 99US-0145192.
PR  23-JUL-1999; 99US-0145145.
PR  23-JUL-1999; 99US-0145216.
PR  26-JUL-1999; 99US-0145224.
PR  27-JUL-1999; 99US-0145276.
PR  27-JUL-1999; 99US-0145913.
PR  27-JUL-1999; 99US-0145918.
PR  27-JUL-1999; 99US-0145919.
PR  28-JUL-1999; 99US-0145951.
PR  02-AUG-1999; 99US-0146386.
PR  02-AUG-1999; 99US-0146388.
PR  02-AUG-1999; 99US-0146389.
PR  03-AUG-1999; 99US-0147038.
PR  04-AUG-1999; 99US-0147204.
PR  04-AUG-1999; 99US-0147302.
PR  05-AUG-1999; 99US-0147192.
PR  05-AUG-1999; 99US-0147260.
PR  06-AUG-1999; 99US-0147303.
PR  06-AUG-1999; 99US-0147416.
PR  09-AUG-1999; 99US-0147493.
PR  09-AUG-1999; 99US-0147935.
PR  10-AUG-1999; 99US-0148171.
PR  11-AUG-1999; 99US-0148319.
PR  12-AUG-1999; 99US-0148341.
PR  13-AUG-1999; 99US-0148565.
PR  13-AUG-1999; 99US-0148684.
PR  16-AUG-1999; 99US-0149368.
```

```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      35.3%; Score 6; DB 21; Length 110;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ILLPDW 8
Db 45 illpdw 50

RESULT 11
AAG61372
ID AAG61372 standard; Protein: 110 AA.
```

```
XX
AC AAG61372;
XX 18-OCT-2000 (first entry)
DT XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 79597.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125786.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
```

```
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-014035.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140693.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142053.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-014433.
PR 19-JUL-1999; 99US-014433.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147410.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159323.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 35.38; Score 6; DB 21; Length 110;
Best Local Similarity 100.08; Pred.No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ILLPDW 8
Db 45 illpdw 50

RESULT 12
AAB41187
ID AAB41187 standard; Protein; 151 AA.
XX
AC AAB41187;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF951 polypeptide sequence SEQ ID NO:1902.
XX
```

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnery; antiparasitic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;

WPI; 2000-602362/57.

N-PSDB; AAC75396.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 1456-1457; 5507pp; English.

AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnery; antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 151 AA;

Query Match 35.3%; Score 6; DB 21; Length 151;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 KEILIE 17

Db 144 keilie 149

RESULT 13

AAG16696

ID AAG16696 standard; Protein; 154 AA.

XX AAG16696;

XX AAG16696;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 17441.

XX Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

XX Arabidopsis thaliana.

XX EF1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149829.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142303.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	09-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144096.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144335.	PR	12-OCT-1999;	99US-0158569.
PR	19-JUL-1999;	99US-0144331.	PR	12-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	20-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161922.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147453.			
PR	09-AUG-1999;	99US-0147955.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			

Query Match

Best Local Similarity

Matches

Score 6;

DB 21;

Length 154;

100.0%;

Pred. No. 51;

Conservative

0;

Mismatches

0;

Indels

0;

Gaps

0;

QY

9 KIRKEI 14

|||||

Db

6 kirkei 11

RESULT 14

AAB93696

ID AAB93696 standard; Protein; 182 AA.

XX

AC AAB93696;


```
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:13267.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
XX
XX Claim 8; SEQ ID 13267; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 182 AA;
XX
XX Query Match 35.3%; Score 6; DB 22; Length 182;
XX Best Local Similarity 100.0%; Pred. No. 59;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 12 KEILIE 17
XX Db 123 keilie 128
XX
XX RESULT 15
XX AAB25223
XX ID AAB25223 standard; Protein; 200 AA.
```

```
XX AAB25223;
XX
XX 27-NOV-2000 (first entry)
XX
XX Eucalyptus grandis cell signalling involved protein SEQ ID NO:542.
XX
XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
XX plant cell signalling; modulation; transgenic plant; pathogen; growth;
XX environmental change; development; cell proliferation; differentiation;
XX elongation; survival; disease resistance; nutrient metabolism.
XX
XX Eucalyptus grandis.
XX WO2000042171-A1.
XX
XX 20-JUL-2000.
XX
XX 11-JAN-2000; 2000WO-US00724.
XX 12-JAN-1999; 99US-0228986.
XX 01-NOV-1999; 99US-0162866.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Strabala TJ, Nieuwenhuizen NJ;
XX WPI; 2000-476052/41.
XX
XX Isolated polynucleotide encoding a polypeptide involved in cell
XX signaling used for generating transgenic plants with modified responses
XX to external signals.
XX
XX Claim 3; Page 241-242; 527pp; English.
XX
XX AA79263 to AA79736 and AAB25100 to AAB25570 represent polynucleotide
XX and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
XX pine (Pinus radiata also known as Monterey pine). The protein sequences
XX are involved in cell signalling. The polynucleotide and protein
XX sequences can be used to modify the response of plant cells to external
XX signals e.g. environmental changes or pathogens during the growth and
XX development of a plant. They can be used to modify cell proliferation,
XX differentiation, elongation and survival, resistance to disease and
XX nutrient metabolism. Examples of modifications which can be produced are
XX altered fruit ripening and senescence of leaves and flowers e.g. to
XX delay senescence and prolong the life of cut flowers or enhance
XX senescence of reproductive organs to engineer sterile plants. Other
XX modifications can be used to delay senescence in selected cell types or
XX organs providing fruit and vegetables which have a longer shelf life
XX between harvest and consumption, or to decrease branching frequency in
XX forest tree species giving long stretches of valuable knot-free clear
XX wood which can be used in solid timber furniture and veneers.
XX
XX Sequence 200 AA;
XX
XX Query Match 35.3%; Score 6; DB 21; Length 200;
XX Best Local Similarity 100.0%; Pred. No. 65;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 4 LLPDWK 9
XX Db 110 llpdwk 115
XX
XX Search completed: January 31, 2002, 13:18:02
XX Job time: 163 sec
```

This Page Blank (uspto)

• • • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:34:55 ; Search time 130.99 seconds
(without alignments)
18.983 Million cell updates/sec

Title: US-08-957-709-11
Perfect score: 17
Sequence: 1 GAILLPDKIRKEILLIE 17

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	7	41.2	1151	4 Q9HCH6	Q9hch6 homo sapien
2	6	35.3	110	10 Q9LNM6	Q9lnm6 arabidopsis
3	6	35.3	132	2 Q9L7U5	Q9l7u5 rhodospiril
4	6	35.3	177	13 Q98U30	Q98u30 limonectes
5	6	35.3	178	2 Q9X0L1	Q9x0l1 thermotoga
6	6	35.3	182	4 Q9NUJ2	Q9nuj2 homo sapien
7	6	35.3	185	1 Q9V1B8	Q9v1b8 pyrococcus
8	6	35.3	223	4 Q75314	Q75314 homo sapien
9	6	35.3	247	2 Q9ZD48	Q9zd48 rickettsia
10	6	35.3	251	1 Q9YAT4	Q9yat4 aeropyrum p
11	6	35.3	259	2 Q9X171	Q9x171 thermotoga
12	6	35.3	262	8 Q21278	Q21278 reclinomona
13	6	35.3	268	1 Q26109	Q26109 methanobact
14	6	35.3	269	1 Q29219	Q29219 archaeoglob
15	6	35.3	272	11 Q70356	Q70356 mus musculu
16	6	35.3	289	2 Q9LAH1	Q9lah1 acetobacter
17	6	35.3	294	5 Q9VB61	Q9vb61 drosophila
18	6	35.3	299	5 Q9XTI1	Q9xti1 caenorhabdi
19	6	35.3	303	10 Q42912	Q42912 malus domes

20 6 35.3 303 10 Q9SX52 09sx52 arabidopsis
21 6 35.3 303 10 Q9LHE7 09lhe7 arabidopsis
22 6 35.3 311 4 O14495 014495 homo sapien
23 6 35.3 320 5 Q9VYT2 09vyt2 drosophila
24 6 35.3 344 10 Q9FGG4 09fgg4 arabidopsis
25 6 35.3 382 2 P72969 p72969 synechocyst
26 6 35.3 387 2 Q9KQW8 09kqm8 vibrio chol
27 6 35.3 391 2 Q9AN62 09an62 bradyrhizob
28 6 35.3 396 4 Q9UHN0 09uhn0 homo sapien
29 6 35.3 410 1 Q9V282 09v282 pyrococcus
30 6 35.3 418 10 Q9FG12 09fg12 arabidopsis
31 6 35.3 508 10 Q9SVA7 09sva7 arabidopsis
32 6 35.3 513 2 O67052 067052 aquifex aeo
33 6 35.3 547 10 Q9SVX6 09svx6 arabidopsis
34 6 35.3 562 2 O84773 084773 chlamydia t
35 6 35.3 564 2 Q9PLF5 09plf5 chlamydia m
36 6 35.3 604 5 O02297 002297 caenorhabdi
37 6 35.3 701 10 Q9FVV6 09fvv6 arabidopsis
38 6 35.3 745 4 Q9Y4Q6 09y4q6 homo sapien
39 6 35.3 781 1 Q9P9N1 09p9n1 sulfurispha
40 6 35.3 863 4 Q9ULH9 09ulh9 homo sapien
41 6 35.3 866 2 O87984 087984 bordetella
42 6 35.3 909 10 Q9ASK1 09ask1 oryza sativ
43 6 35.3 992 2 Q9AJ63 09aj63 rickettsia
44 6 35.3 1005 2 Q9AJ36 09aj36 rickettsia
45 6 35.3 1337 2 P76465 09aj36 rickettsia
P76465 escherichia

ALIGNMENTS

RESULT 1
Q9HCH6 ID Q9HCH6 PRELIMINARY; PRT; 1151 AA.
AC Q9HCH6
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE K1AA1596 PROTEIN (FRAGMENT).
GN K1AA1596.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
XVIII. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro."
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046816; BAB13422.1;
FT NON_TER 1
SQ SEQUENCE 1151 AA; 130000 MW; ED6120D527699C04 CRC64;

Query Match 41.2%; Score 7; DB 4; Length 1151;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KIRKEIL 15
Db 455 KIRKEIL 461

RESULT 2
Q9LNM6 ID Q9LNM6 PRELIMINARY; PRT; 110 AA.
AC Q9LNM6
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE K1AA1596 PROTEIN (FRAGMENT).
GN K1AA1596.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
XVIII. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro."
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046816; BAB13422.1;
FT NON_TER 1
SQ SEQUENCE 1151 AA; 130000 MW; ED6120D527699C04 CRC64;

DE F12K21.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
 RA Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F12K21 from chromosome
 RT I";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E.,
 RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC03279; AAF79252.1;
 DR EMBL; AC03279; AAF79252.1;
 SQ SEQUENCE 110 AA; 12686 MW; 343CEEC1816A74D5 CRC64;

Query Match 35.3%; Score 6; DB 10; Length 110;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 ILLPDW 8
 Db 45 ILLPDW 50

RESULT 3
 O9L7U5 PRELIMINARY; PRT; 132 AA.
 ID O9L7U5
 DT 01-OCT-2000 (TREMREL. 15, Created)
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
 DT 01-OCT-2000 (TREMREL. 15, Last annotation update)
 DE FLAGELLAR M RING PROTEIN FLIF (FRAGMENT).
 GN FLIF.
 OS Rhodospirillum rubrum (Rhodocista centenario).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 OC Rhodospirillum.
 OX NCBI_TaxID=34018;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McClean J.R., Rollo D., Bauer C.E., Rushing B.R., Jiang Z.;
 RT "Gene duplication and lateral transfer events giving rise to
 RT Rhodospirillum rubrum polar and lateral flagellar motor switch
 RT components";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF220002; AAF35844.1;
 FT NON_TER 132
 SQ SEQUENCE 132 AA; 13601 MW; E5E0C543E479CD1 CRC64;

Query Match 35.3%; Score 6; DB 2; Length 132;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAIILP 6
 Db 59 CAIILP 64

RESULT 4
 O98U30 PRELIMINARY; PRT; 177 AA.
 ID O98U30
 DT 01-JUN-2001 (TREMREL. 17, Created)
 DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE TYROSINASE PRECURSOR (FRAGMENT).
 OS Limnodynastes kuhlii.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae;
 OC Limnodynastes.
 OX NCBI_TaxID=110107;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bessuyt F., Milinkovitch M.C.;
 RT "Revealing Adaptive Radiations in Madagascar and Asian Ranid Frogs
 RT Reveal Co-variation between Larval and Adult Traits";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF249179; AAG49822.1;
 FT NON_TER 177
 SQ SEQUENCE 177 AA; 20423 MW; 28A72B222A031854 CRC64;

Query Match 35.3%; Score 6; DB 13; Length 177;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 IRKEIL 15
 Db 61 IRKEIL 66

RESULT 5
 O9X0L1 PRELIMINARY; PRT; 178 AA.
 ID O9X0L1
 DT 01-NOV-1999 (TREMREL. 12, Created)
 DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE HYDROTICAL 19.9 KDA PROTEIN.
 GN TM127.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton M.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima";
 RL Nature 399:323-329(1999).
 DR EMBL; AE001770; AAD36203.1;
 DR TIGR; TM127;
 KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 178 AA; 19894 MW; F6665DA03B4F4C0D CRC64;

Query Match 35.3%; Score 6; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 IRKEIL 15
Db 69 IRKEIL 74

RESULT 6

ID Q9NUJ2 PRELIMINARY; PRT; 182 AA.

AC Q9NUJ2;

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DE CDNA FLJ11339 FIS, CLONE PLACE1010743, WEAKLY SIMILAR TO HOMO SAPIENS

DE MYOSIN-IXB SPLICE VARIANT MRNA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,

RA Tanai H., Kinata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,

RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,

RA Masuho Y., Kanehori K.

RT "NEDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBSJ databases.

DR EMBL; AK002201; BAA92132.1; -.

SQ SEQUENCE 182 AA; 20362 MW; 4311D85F0B587E20 CRC64;

Query Match 35.3%; Score 6; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 KEILIE 17
Db 123 KEILIE 128

RESULT 7

ID Q9VIB8 PRELIMINARY; PRT; 185 AA.

AC Q9VIB8;

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE 2-KETOGLUTARATE FERREDOXIN OXIDOREDUCTASE, SUBUNIT GAMMA (KORG-2).

GN KORG-2 OR PAB0348.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI_TaxID=29292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORSAY;

RA Heilig R.

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBSJ databases.

DR EMBL; AJ248284; CAB49431.1; -.

DR InterPro; IPR002869; POR.

DR Pfam; PF01558; POR; 1.

KW Complete proteome.

SQ SEQUENCE 185 AA; 20022 MW; 41D78318F6C7523D CRC64;

Query Match 35.3%; Score 6; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RKEILI 16
Db 5 RKEILI 10

RESULT 8

ID O75314 PRELIMINARY; PRT; 223 AA.

AC O75314;

DT 01-NOV-1998 (TRENBLrel. 08, Created)

DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)

DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)

DE MYOSIN-IXB SPLICE VARIANT (FRAGMENT).

GN MYO9B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97063843; PubMed=8907710;

RA Wirth J.A., Jensen K.A., Post P.L., Bement W.M., Mooseker M.S.;

RT "Human myosin-IXb, an unconventional myosin with a chimerin-like rho/rac GTPase-activating protein domain in its tail.";

RL J. Cell Sci. 109:0-0(0).

[2]

RP SEQUENCE FROM N.A.

RX MEDLINE=98158729; PubMed=9490638;

RA Post P., Bokoch G., Mooseker M.;

RT "Human myosin-IXb is a mechanotchemically active motor and a GAP for rho.";

RL J. Cell Sci. 111:0-0(0).

DR EMBL; AF020267; AAC26597.1; -.

FT NON_TER 1

SQ SEQUENCE 223 AA; 24494 MW; E474CDDE07810EBB CRC64;

Query Match 35.3%; Score 6; DB 4; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 KEILIE 17
Db 25 KEILIE 30

RESULT 9

ID Q9ZD48 PRELIMINARY; PRT; 247 AA.

AC Q9ZD48;

DT 01-MAY-1999 (TRENBLrel. 10, Created)

DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)

DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)

DE HYPOTHETICAL 28.2 KDA PROTEIN.

GN RP499.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=782;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MADRID E;

RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,

RA Sierichitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of mitochondria.";

```

RL Nature 396:133-140(1998).
DR EMBL; AJ235272; CAAL4951.1; -.
SW Hypothetical protein; Complete proteome.
SQ SEQUENCE 247 AA; 28210 MW; 606D03AC2B030392 CRC64;

Query Match 35.3%; Score 6; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 IRKEIL 15
    |||||
Db 80 IRKEIL 85

RESULT 10
ID Q9YAT4 PRELIMINARY; PRT; 251 AA.
AC Q9YAT4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHEICAL 26.7 KDA PROTEIN APEL860.
GN APEL860.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
RA Hosoyma A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80864.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 251 AA; 26654 MW; 08BCC7A0C471EC64 CRC64;

Query Match 35.3%; Score 6; DB 1; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAILLP 6
    |||||
Db 123 GAILLP 128

RESULT 11
ID Q9X171 PRELIMINARY; PRT; 259 AA.
AC Q9X171;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LIPASE, PUTATIVE.
GN TMI350.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,

```

```

RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -I- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DR EMBL; AE001789; AAD36421.1; -.
DR TIGR; TMI350; -.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR003089; AB_hydrolase.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR Pfam; PF00561; abhydrolase_1.
DR PRINTS; PR00111; ABHYDROLASE.
DR Complete proteome.
SQ SEQUENCE 259 AA; 29603 MW; 426CCE1F16A9EC97 CRC64;

Query Match 35.3%; Score 6; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AILLPD 7
    |||||
Db 113 AILLPD 118

RESULT 12
ID O21278 PRELIMINARY; PRT; 262 AA.
AC O21278;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SUBUNIT OF ABC TRANSPORTER FOR CYTOCHROME C1.
GN YEJU.
OS Reclinomonas americana.
OC Mitochondrion.
OC Eukaryota; Core jakobids; Reclinomonas.
OX NCBI_TaxID=48483;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50394;
RX MEDLINE=97311393; PubMed=9168110;
RA Lang B.F., Burger G., O'Kelly C.J., Cedergren R., Golding G.B.,
RA Lemieux C., Sankoff D., Turmel M., Gray M.W.;
RT "An ancestral mitochondrial DNA resembling a eubacterial genome in
RT miniature."
RL Nature 387:493-497(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50394;
RA Lang B.F., Burger G.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007261; AAD11905.1; -.
DR InterPro; IPR002541; cytc_asm.
DR InterPro; IPR003557; cytc_biolg_ccmc.
DR Pfam; PF01578; Cytc_asm; 1.
DR PRINTS; PR01386; CCMCBIOGNIS.
DR Mitochondrion.
SQ SEQUENCE 262 AA; 30542 MW; 69623E67457547C5 CRC64;

Query Match 35.3%; Score 6; DB 8; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RKEILI 16
    |||||
Db 237 RKEILI 242

```

```
RESULT 13
ID O26109 PRELIMINARY; PRT; 268 AA.
AC O26109;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CONSERVED PROTEIN.
GN MTH1.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., DuBois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RA *Complete genome sequence of Methanobacterium thermoautotrophicum
RA delah: functional analysis and comparative genomics.*;
RT J. Bacteriol. 179:7135-7155(1997).
RL EMBL: AE000795; AAB84510.1;
DR InterPro: IPR003750; DUF171.
DR Pfam: PF02598; DUF171; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 268 AA; 30191 MW; 1B97DB1CEE27BAED CRC64;

Query Match 35.3%; Score 6; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 KEILIE 17
DB 149 KEILIE 154

RESULT 14
ID O29219 PRELIMINARY; PRT; 269 AA.
AC O29219;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 30.7 KDA PROTEIN.
GN AF1043.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervilavge A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
```

```
RL Nature 390:364-370(1997).
DR EMBL: AE001031; AAB90208.1;
DR TIGR: AF1043;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 269 AA; 30670 MW; 6A41E02B25FBC0E0 CRC64;

Query Match 35.3%; Score 6; DB 1; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 KIRKEI 14
DB 245 KIRKEI 250

RESULT 15
O70356 PRELIMINARY; PRT; 272 AA.
AC O70356;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BUTYROPHILIN-LIKE (FRAGMENT).
GN BTNL2 OR NG10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Qin S., Loretz C., Mix L., Lasky S., Madan A., Hood L.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL: AF050157; AAC05289.1;
DR MGD; MGI:1859549; Btln2.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig; 1.
DR SMART; SMO0406; IGV; 1.
DR NON_TER 272
FT SEQUENCE 272 AA; 30334 MW; E745E84220343663 CRC64;

Query Match 35.3%; Score 6; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AILLPD 7
DB 234 AILLPD 239

Search completed: January 31, 2002, 13:37:30
Job time: 155 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:38:00 ; Search time 46.78 Seconds
(without alignments)
13.324 Million cell updates/sec

Title: US-08-957-709-11

Perfect score: 17
Sequence: 1 GAILLPDWKIRKEILLIE 17Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	14	82.4	156	1 DCD_PYRHO	O57706 pyrococcus
2	13	76.5	154	1 DCD_PYRAB	Q9uxs8 pyrococcus
3	6	35.3	202	1 SODM_STRMU	P09738 streptococc
4	6	35.3	220	1 Y132_METJA	O57596 methanococc
5	6	35.3	249	1 Y768_METJA	O58178 methanococc
6	6	35.3	260	1 YLBK_BACSU	O34731 bacillus su
7	6	35.3	284	1 CELA_ACEXY	P27897 acetobacter
8	6	35.3	285	1 YFIE_BACSU	P54721 bacillus su
9	6	35.3	336	1 Y883_METJA	O58293 methanococc
10	6	35.3	344	1 STSY_RAUSE	P15324 rauevolfia s
11	6	35.3	478	1 HLYD_PASHA	P16534 pasteurella
12	6	35.3	478	1 HLYD_PASSP	P55125 pasteurella
13	6	35.3	583	1 T2F1_FLAOK	P14870 flavobacter
14	6	35.3	595	1 P2X7_RAT	Q64663 rattus norv
15	6	35.3	799	1 YJVT_YEAST	P40897 saccharomyc
16	6	35.3	829	1 GUTR_BACSU	P39143 bacillus su
17	6	35.3	854	1 CLPC_CHLTR	O84288 chlamydia t
18	6	35.3	870	1 CLPC_CHLMU	O9pkas chlamydia m
19	6	35.3	992	1 SCAM_RICTY	O9aj63 rickettsia
20	6	35.3	1022	1 SCAM_RICPR	Q9zd49 rickettsia
21	6	35.3	1181	1 Y102_METJA	O60301 methanococc
22	6	35.3	1391	1 N157_YEAST	P40064 saccharomyc
23	6	35.3	1756	1 PEPL_HUMAN	O60437 homo sapien
24	6	35.3	2032	1 CTOG_HUMAN	O14008 homo sapien
25	6	35.3	2158	1 MY9B_HUMAN	Q13459 homo sapien
26	5	29.4	52	1 Y567_PSEAE	O915w9 pseudomonas
27	5	29.4	54	1 RC2A_SYNY3	O9znq7 arabidopsis
28	5	29.4	54	1 YB69_SYNY3	P74805 synecocyst
29	5	29.4	58	1 VPU_HV12H	P08806 human immun
30	5	29.4	73	1 RC23_ARATH	O9m095 arabidopsis
31	5	29.4	74	1 RC24_ARATH	O9su10 arabidopsis
32	5	29.4	75	1 RC22_ARATH	O82232 arabidopsis
33	5	29.4	92	1 FERN_AZOVI	P11054 azotobacter

34 5 29.4 102 1 CH10_SYNVU O50322 synecococc
35 5 29.4 103 1 CH10_SYNY3 Q05971 synecocyst
36 5 29.4 103 1 CH10_CVAPA Q37761 cyanophora
37 5 29.4 112 1 YAFU_ECOLI P77354 escherichia
38 5 29.4 113 1 YNFB_ECOLI P76170 escherichia
39 5 29.4 115 1 NU3M_ALLMI O47874 alligator m
40 5 29.4 116 1 NU3M_ORNAN Q36456 ornithorhyn
41 5 29.4 116 1 YCFF_HAEIN P44956 haemophilus
42 5 29.4 125 1 HINT_BOVIN P16436 bos taurus
43 5 29.4 125 1 HINT_HUMAN P49773 homo sapien
44 5 29.4 125 1 HINT_MOUSE P70349 mus musculus
45 5 29.4 125 1 HINT_RABIT P80912 oryctolagus

ALIGNMENTS

RESULT 1
ID DCD_PYRHO STANDARD; PRT; 156 AA.
AC O57706;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE).
GN DCD OR PH1997.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Negai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -|- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -|- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000007; BAA31124.1;
DR InterPro; IPR003232; DCTP_deaminase.
DR InterPro; IPR001428; dutpase.
DR Pfam; PF00692; dutpase; 1.
DR ProDom; PD004900; dctp_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 156 AA; 17871 MW; 96B2C2C50393A985 CRC64;

Query Match 82.4%; Score 14; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 4.1e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLPDWKIRKEILLIE 17

Db 2 LLPDWKIRKEILLIE 15

RESULT 2

```

DCD_PYRAB
ID DCD_PYRAB STANDARD; PRT; 154 AA.
AC Q9UXS8;
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE).
GN DCD OR PAB1164.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution." the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O -> DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248288; CAB50685.1; -
DR InterPro; IPR003232; DCTP_deaminase.
DR Pfam; PF00692; DUTPase; 1.
DR ProDom; PD004900; DCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 154 AA; 17758 MW; C031BEE419094DDB CRC64;

Query Match 76.5%; Score 13; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 4.6e-07;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LLPDWKIRKEILI 16
Db 2 LLPDWKIRKEILI 14

RESULT 3
SODM_STRMU STANDARD; PRT; 202 AA.
AC P09736; Q59791;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SUPEROXIDE DISMUTASE [MN-FE] (EC 1.15.1.1).
GN SOD A OR SOD.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS-5;
RA MEDLINE=92332426; PubMed=1321118;
RA Nakayama K.;
RT "Nucleotide sequence of Streptococcus mutans superoxide dismutase gene and isolation of insertion mutants.";
RL J. Bacteriol. 174:4928-4934(1992).
RN [2]
RP SEQUENCE OF 1-22.
RX MEDLINE=86250886; PubMed=3722201;
RA Martin M.E., Byers B.R., Olson M.O.J., Salin M.L., Arceneaux J.E.L.,

```

```

RA Tolbert C.;
RT "A Streptococcus mutans superoxide dismutase that is active with either manganese or iron as a cofactor.";
RL J. Biol. Chem. 261:9361-9367(1986).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).
CC -1- COFACTOR: ACTIVE WITH EITHER MANGANESE OR IRON.
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S39782; AAB22503.1; -
DR PIR; A24812; A24812.
DR HSP; P09214; IHMG.
DR InterPro; IPR001189; SOD_M1.
DR Pfam; PF00081; sodfe; 1.
DR ProDom; PD000475; SOD_M1; 1.
DR PROSITE; PS00088; SOD_MW; 1.
KW Oxidoreductase; Manganese; Iron.
FT INIT_MET 0
FT METAL 26 26 MANGANESE (BY SIMILARITY).
FT METAL 80 80 MANGANESE (BY SIMILARITY).
FT METAL 162 162 MANGANESE (BY SIMILARITY).
FT METAL 166 166 MANGANESE (BY SIMILARITY).
FT CONFLICT 3 3 L -> T (IN REF. 2).
SQ SEQUENCE 202 AA; 22494 MW; C0C853BE0032B541 CRC64;

Query Match 35.3%; Score 6; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AILLPD 7
Db 1 AILLPD 6

RESULT 4
Y132_METJA STANDARD; PRT; 220 AA.
AC Q57536; 1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0132.
GN MJ0132.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D., Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I., Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A., Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D., Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C., Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M., Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii";
RL Science 273:1058-1073(1996).

```

CC CC -1- SIMILARITY: TO M.JANNASCHII MJ1220 AND MJEC142.
 CC CC -1- SIMILARITY: WITH TYPE I RESTRICTION SYSTEM ADENINE METHYLASES
 CC CC (M SUBUNIT).
 CC CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 CC DR EMBL; U67470; AAB98113.1; -;
 CC DR TIGR; MJ0132; -;
 CC DR InterPro; IPR003356; N6_DNA_Mtase.
 CC DR Pfam; PF02384; N6_Mtase; 1.
 CC DR Hypothetical protein; Complete proteome.
 CC KW SEQUENCE 220 AA; 25766 MW; 710DDAE4C7A47954 CRC64;
 CC SO
 CC
 CC Query Match 35.3%; Score 6; DB 1; Length 220;
 CC Best Local Similarity 100.0%; Pred. No. 14;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 9 KIRKEI 14
 CC DB 66 KIRKEI 71
 CC
 CC RESULT 5
 CC ID Y768_METJA STANDARD; PRT; 249 AA.
 CC AC Q58178;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
 CC DE HYPOTHETICAL PROTEIN MJ0768.
 CC GN MJ0768.
 CC OS Methanococcus jannaschii.
 CC OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC OC Methanococcus
 CC OX NCBI_TaxID=2190;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 CC RX MEDLINE=96337999; PubMed=8688087;
 CC RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 CC RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 CC RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 CC RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 CC RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 CC RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 CC RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 CC RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 CC RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 CC jannaschii".
 CC RL Science 273:1058-1073(1996).
 CC CC -1- SIMILARITY: TO THE N-TERMINAL OF M.TUBERCULOSIS MTCY71.02.
 CC CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 CC DR EMBL; U67522; AAB98763.1; -;
 CC DR TIGR; MJ0768; -;
 CC DR InterPro; IPR002847; DUF129.
 CC DR Pfam; PF01996; DUF129; 1.
 CC KW Hypothetical protein; Complete proteome.
 CC SO SEQUENCE 249 AA; 27147 MW; A6B66FF4B52C15A2 CRC64;

Query Match 35.3%; Score 6; DB 1; Length 249;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 KIRKEI 14
 DB 136 KIRKEI 141
 RESULT 6
 YLBK_BACSU
 ID YLBK_BACSU STANDARD; PRT; 260 AA.
 AC O34731;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 28.3 KDA PROTEIN IN CTAG-RPME INTERGENIC REGION.
 GN YLBK.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Bartero M., Presecan E., Glaser P., Richou A., Danchin A.;
 RT "Bacillus subtilis chromosomal region downstream nprE";
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0028 (SWS) FAMILY.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 CC DR EMBL; Z98682; CAB11357.1; -;
 CC DR EMBL; Z99111; CAB13377.1; -;
 CC DR Subtilist; BG13363; Ylbk.
 CC DR InterPro; IPR001423; UPF0028.
 CC DR Pfam; PF01173; UPF0028; 1.
 CC DR POSITE; PS01237; UPF0028; 1.
 CC KW Hypothetical protein; Complete proteome.
 CC SQ SEQUENCE 260 AA; 28258 MW; 363B6DDF017BE77E CRC64;
 CC
 CC Query Match 35.3%; Score 6; DB 1; Length 260;
 CC Best Local Similarity 100.0%; Pred. No. 16;
 CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 9 KIRKEI 14
 CC DB 249 KIRKEI 254
 RESULT 7
 CELA_ACEXY
 ID CELA_ACEXY STANDARD; PRT; 284 AA.
 AC P27897;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE UTP--GLUCOSE-1-PHOSPHATE URIDYLTRANSFERASE (EC 2.7.7.9) (UDP-GLUCOSE
 DE PYROPHOSPHORYLASE) (UDPGP) (ALPHA-D-GLUCOSYL-1-PHOSPHATE
 DE URIDYLTRANSFERASE) (URIDINE DIPHOSPHOGLUCOSE PYROPHOSPHORYLASE).
 GN CELA.
 OS Acetobacter xylinum.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Gluconacetobacter.

```

OX NCBI_TaxID=28448;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=92041596; PubMed=1938907;
RX Brede G., Fjaervik E., Valla S.;
RT "Nucleotide sequence and expression analysis of the Acetobacter
RT xylX gene, a 12 kb nucleotide sequence around the 76 degrees
RL J. Bacteriol. 173:7042-7045(1991).
CC -1- CATALYTIC ACTIVITY: UDP + ALPHA-D-GLUCOSE 1-PHOSPHATE -
CC PYROPHOSPHATE + UDP-GLUCOSE.
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC UDPGP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M7548; AAA21888.1; -
DR PIR; A41382;
DR InterPro: IPR001825; NTP_transferase.
DR Pfam; PF00483; NTP_transferase; 1.
KW Transferase; Kinase; Nucleotidyltransferase.
SQ SEQUENCE 284 AA; 30992 MW; 6C809AFFBBA1791B CRC64;

Query Match 35.3%; Score 6; DB 1; Length 284;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AILLPD 7
DB 129 AILLPD 134

RESULT 8
YFIE_BACSU
ID YFIE_BACSU STANDARD; PRT; 285 AA.
AC P54721;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 31.5 KDA PROTEIN IN GLVBC 3'REGION.
GN YFIE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96262713; PubMed=8704981;
RX Yamamoto H., Uchiyama S., Fajar A.N., Ogasawara N., Sekiguchi J.;
RT "Determination of a 12 kb nucleotide sequence around the 76 degrees
RT region of the bacillus subtilis chromosome."
RL Microbiology 142:1417-1421(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; D50543; BAA09109.1; -
DR EMBL; Z99108; CAB12653.1; -
DR Subtilist; BG11852; YfIE.
DR Pfam; PF00903; Glyoxalase; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 285 AA; 31521 MW; 79E4F242334825E2 CRC64;

NCBI_TaxID=28448;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=92041596; PubMed=1938907;
RX Brede G., Fjaervik E., Valla S.;
RT "Nucleotide sequence and expression analysis of the Acetobacter
RT xylX gene, a 12 kb nucleotide sequence around the 76 degrees
RL J. Bacteriol. 173:7042-7045(1991).
CC -1- CATALYTIC ACTIVITY: UDP + ALPHA-D-GLUCOSE 1-PHOSPHATE -
CC PYROPHOSPHATE + UDP-GLUCOSE.
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC UDPGP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M7548; AAA21888.1; -
DR PIR; A41382;
DR InterPro: IPR001825; NTP_transferase.
DR Pfam; PF00483; NTP_transferase; 1.
KW Transferase; Kinase; Nucleotidyltransferase.
SQ SEQUENCE 284 AA; 30992 MW; 6C809AFFBBA1791B CRC64;

Query Match 35.3%; Score 6; DB 1; Length 285;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AILLPD 7
DB 77 AILLPD 82

RESULT 9
Y883_METJA
ID Y883_METJA STANDARD; PRT; 336 AA.
AC Q88233; 1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0883.
GN MJ0883.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RX Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RX Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RX Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RX Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RX Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RX Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RX Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RX Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO M.JANNASCHII M1557 AND TO YEAST YHR070W.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U67532; AAB98887.1; -
DR TIGR; MJ0883; -
DR InterPro: IPR003402; Met_L0.
DR InterPro: IPR000051; SAM_bind.
DR Pfam; PF02475; Met_L0; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 336 AA; 39000 MW; 3E4811F0932EE95F CRC64;

Query Match 35.3%; Score 6; DB 1; Length 336;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KIRKEI 14
DB 115 KIRKEI 120

RESULT 10
STSY_RAUSE
ID STSY_RAUSE STANDARD; PRT; 344 AA.
AC P15324;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)

```

15-JUL-1998 (Rel. 36, Last annotation update)
 DT STRICTOSIDINE SYNTHASE PRECURSOR (EC 4.3.3.2).
 GN STRI.
 OS Rauvolfia serpentina (serpentwood) (Devilpepper), and
 OS Rauvolfia mannii
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
 OC Vinceae; Rauvolfia.
 OX NCBI_TaxID=4060, 4062;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC SPECIES=R. serpentina;
 RA MEDLINE=89005637; PubMed=3049153;
 RX Kutchan T.M., Hamp N., Lottspeich F., Beyreuther K., Zenk M.H.;
 RT "The cDNA clone for strictosidine synthase from Rauvolfia serpentina.
 RL DNA sequence determination and expression in Escherichia coli.";
 FEBS Lett. 237:40-44(1988).
 [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=R. serpentina, and R. mannii;
 RX MEDLINE=92231602; PubMed=1567228;
 RA Bracher D., Kutchan T.M.;
 RT "Strictosidine synthase from Rauvolfia serpentina: analysis of a gene
 involved in indole alkaloid biosynthesis.";
 RL Arch. Biochem. Biophys. 294:717-723(1992).
 CC -1- FUNCTION: CATALYZES THE STEREOSPECIFIC CONDENSATION OF
 CC TRYPTAMINE WITH SCOLOGANIN TO FORM STRICTOSIDINE, THE KEY
 CC INTERMEDIATE OF INDOLE ALKALOID BIOSYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: 3-ALPHA(S)-STRICTOSIDINE = TRYPTAMINE +
 CC SCOLOGANIN.
 CC -1- PATHWAY: ALKALOID BIOSYNTHESIS.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: VACUOLAR.
 CC -1- SIMILARITY: BELONGS TO THE STRICTOSIDINE SYNTHASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Y00756; CAA68725.1; -;
 DR EMBL; X62334; CAA44208.1; -;
 DR EMBL; X63431; CAA45025.1; -;
 DR PIR; S01325; S01325.
 DR PIR; S21868; S21868.
 DR PIR; S21718; S21718.
 DR PIR; S21718; S21718.
 KW Lyase; Glycoprotein; Alkaloid metabolism; Signal.
 FT SIGNAL 1 22 BY SIMILARITY.
 FT CHAIN 23 344 STRICTOSIDINE SYNTHASE.
 FT CARBOHYD 91 91 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SEQUENCE 344 AA; 38162 MW; 7CD38882621F768A CRC64;
 Query Match 35.3%; Score 6; DB 1; Length 344;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 KEILIE 17
 Db 32 KEILIE 37
 RESULT 11
 HLYD_PASHA STANDARD; PRT; 478 AA.
 AC P16534;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

LEUKOTOXIN SECRETION PROTEIN D.
 GN LKTD.
 OS Pasteurella haemolytica.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Mannheimia.
 OX NCBI_TaxID=75985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROTYPE A1;
 RX MEDLINE=87306837; PubMed=3040588;
 RA Lo R.Y.C., Strathdee C.A., Shewen P.E.;
 RT "Nucleotide sequence of the leukotoxin genes of Pasteurella
 RL haemolytica A1";
 Infect. Immun. 55:1987-1996(1987).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROTYPE A1;
 RX MEDLINE=89123172; PubMed=2914876;
 RA Strathdee C.A., Lo R.Y.C.;
 RT "Cloning, nucleotide sequence, and characterization of genes encoding
 RL the secretion function of the Pasteurella haemolytica leukotoxin
 RT determinant";
 J. Bacteriol. 171:916-928(1989).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROTYPE A1 / PHL101;
 RX MEDLINE=89210283; PubMed=2707120;
 RA Highlander S.K., Chidambaram M., Engler M.J., Weinstock G.M.;
 RT "DNA sequence of the Pasteurella haemolytica leukotoxin gene
 RL cluster";
 DNA 8:15-28(1989).
 CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF THE LEUKOTOXIN.
 CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-BOUND (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M20730; AAA25531.1; -;
 DR EMBL; M24197; AAA25545.1; -;
 DR PIR; B32051; B32051.
 DR PIR; S29518; S29518.
 DR InterPro; IPR002215; HlyD.
 DR Pfam; PF00529; HlyD; 1.
 DR PROSITE; PS00543; HLYD_FAMILY; 1.
 KW Hemolysis; Cytolysis; Transport; Transmembrane; Inner membrane.
 FT DOMAIN 1 59 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 60 80 POTENTIAL.
 FT DOMAIN 81 478 PERIPLASMIC (POTENTIAL).
 FT CONFLICT 18 18 T -> I (IN REF. 3).
 FT CONFLICT 47 47 D -> E (IN REF. 3).
 FT CONFLICT 150 150 N -> T (IN REF. 3).
 FT CONFLICT 236 236 F -> L (IN REF. 3).
 FT CONFLICT 266 267 EL -> AV (IN REF. 3).
 FT SEQUENCE 478 AA; 54761 MW; E2B8B5FF5AF988C1 CRC64;
 Query Match 35.3%; Score 6; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 WKIRKE 13
 Db 23 WKIRKE 28
 RESULT 12
 HLYD_PASS

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M33;
RX MEDLINE=96202420; PubMed=8614837;
RA Surprenant A., Rassendren F., Kawashima E., North R.A., Buell G.N.;
RT "The cytolytic P2X receptor for extracellular ATP identified as a P2X
receptor (P2X7).";
RL Science 272:735-738(1996).
CC -1- FUNCTION: RECEPTOR FOR ATP THAT ACTS AS A LIGAND GATED ION
CHANNEL. RESPONSIBLE FOR ATP-DEPENDENT LYSIS OF MACROPHAGES
THROUGH THE FORMATION OF MEMBRANE PORES PERMEABLE TO LARGE
MOLECULES. COULD FUNCTION IN BOTH FAST SYNAPTIC TRANSMISSION AND
THE ATP-MEDIATED LYSIS OF ANTIGEN-PRESENTING CELLS.
CC -1- SUBUNIT: HOMO- OR HETEROPOLYMERS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE P2X RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X95882; CAA65131.1; -;
DR InterPro; IPR001429; P2X_receptor.
DR Pfam; PF00864; P2X_receptor; 1.
DR PRINTS; PR01307; P2XRECEPTOR.
DR PROSITE; PRO1314; P2X7RECEPTOR.
DR PROSITE; PS01212; P2X_RECEPTOR; 1.
KW Ionic channel; Transmembrane; Ion transport; Receptor; Glycoprotein.
FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 26 46 1 (POTENTIAL).
FT DOMAIN 47 334 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 335 355 2 (POTENTIAL).
FT DOMAIN 356 595 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 595 AA; 68391 MW; 4A6DD6058E5988D3 CRC64;

Query Match 35.3%; Score 6; DB 1; Length 595;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 WKIRKE 13
Db 575 WKIRKE 580
|||||
RESULT 15
YJV2_YEAST STANDARD; PRT; 799 AA.
AC P40897;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 91.6 KDA PROTEIN IN HXT8-CRT1 INTERGENIC REGION.
GN YJL212C OR J0236 OR HRD799.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=S288C;
RX MEDLINE=95242842; PubMed=7725802;
RA Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D.,
RA Hilger F.;
RT "Sequence analysis of a 40.2 kb DNA fragment located near the left
telomere of yeast chromosome X.";
RL Yeast 10:1657-1662(1994).
CC -1- SIMILARITY: STRONG, TO S.POMBE ISP4.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z34098; CAA83999.1; -;
DR EMBL; Z49487; CAA89509.1; -;
DR PIR; S45161; S45161.
DR SGD; S0003748; YJL212C.
KW Hypothetical protein.
SQ SEQUENCE 799 AA; 91615 MW; AF81676CC9B30759 CRC64;

Query Match 35.3%; Score 6; DB 1; Length 799;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LLPDWK 9
Db 154 LLPDWK 159
|||||
Search completed: January 31, 2002, 13:39:10
Job time: 70 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:19 ; Search time 78.64 Seconds
(without alignments)
16.467 Million cell updates/sec

Title: US-08-957-709-11

Perfect score: 17
Sequence: 1 GAILLPDWKIRKEILIE 17

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	82.4	156	2 E71216	dCTP deaminase (EC
2	13	76.5	154	2 G75030	dCTP deaminase (EC
3	6	35.3	178	2 D72293	hypothetical prote
4	6	35.3	185	2 H75168	2-ketoglutarate fe
5	6	35.3	203	2 A42710	superoxide dismuta
6	6	35.3	220	2 D64316	restriction modifi
7	6	35.3	247	2 E71653	hypothetical prote
8	6	35.3	249	1 H64395	hypothetical prote
9	6	35.3	251	2 C72572	hypothetical prote
10	6	35.3	259	2 F72264	hypothetical prote
11	6	35.3	260	2 H69874	conserved hypotet
12	6	35.3	262	2 S78172	probable transport
13	6	35.3	268	1 A69000	conserved hypotet
14	6	35.3	269	2 C69380	hypothetical prote
15	6	35.3	284	1 A41382	UMP--glucose-1-pho
16	6	35.3	285	2 H69802	conserved hypotet
17	6	35.3	299	2 T20141	hypothetical prote
18	6	35.3	303	2 H96539	hypothetical prote
19	6	35.3	303	2 T17012	probable phosphop
20	6	35.3	336	2 C64410	hypothetical prote
21	6	35.3	342	2 S29894	strictosidine synt
22	6	35.3	344	2 S01325	hypothetical prote
23	6	35.3	382	2 S74947	hypothetical prote
24	6	35.3	387	2 D82134	benzoate transport
25	6	35.3	410	2 E75208	probable valine--p
26	6	35.3	478	2 D30169	leukotoxin secreti
27	6	35.3	508	2 T09367	cytochrome P450 ho
28	6	35.3	513	2 A70378	histidine kinase s
29	6	35.3	547	2 T06758	probable galactose

30	6	35.3	562	2 C71473	hypothetical prote
31	6	35.3	564	2 E81735	conserved hypotet
32	6	35.3	583	2 A32861	type II site-speci
33	6	35.3	604	2 T24456	hypothetical prote
34	6	35.3	701	2 G96737	ARE1-like protein,
35	6	35.3	739	2 T49456	hypothetical prote
36	6	35.3	745	2 T12528	hypothetical prote
37	6	35.3	781	2 JC7382	DNA-directed DNA p
38	6	35.3	799	2 S50773	probable membrane
39	6	35.3	829	2 T40014	sorbitol dehydroge
40	6	35.3	854	1 C71533	endopeptidase Clp
41	6	35.3	870	2 B81689	ATP-dependent Clp
42	6	35.3	1181	2 B64516	hypothetical prote
43	6	35.3	1337	2 B64993	hypothetical prote
44	6	35.3	1391	2 S50608	hypothetical prote
45	6	35.3	1426	2 T00337	hypothetical prote

ALIGNMENTS

RESULT 1

E71216

dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: E71216

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic A:Reference number: A71000; MUID:98344137

A:Accession: E71216

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-156 <KAW>

A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31124.1; PID:g3258441

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa C:Genetics:

A:Gene: PH1997

C:Superfamily: dCTP deaminase

C:Keywords: hydrolase

Query Match 82.4%; Score 14; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 LLPDWKIRKEILIE 17

Db 2 LLPDWKIRKEILIE 15

RESULT 2

G75030

dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: G75030

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: G75030

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-154 <KAW>

A:Cross-references: GB:AJ248289; GB:AL096836; NID:g5458960; PIDN:CAB50685.1; PID:g545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: dcd; PAB1164

C:Superfamily: dCTP deaminase

C;Keywords: hydrolase

Query Match 76.5%; Score 13; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 1; 1e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 LLPDKIRKEILI 16
|||||
Db 2 LLPDKIRKEILI 14

RESULT 3

D72293
hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 28-Jul-2000
C:Accession: D72293
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
A:Accession: D72293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-178 <KAW>
A:Cross-references: GB:AE001770; GB:AE000512; NID:94981658; PIDN:AAD36203.1; PID:9498167
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1127
C:Superfamily: Thermotoga maritima hypothetical protein TM1127

Query Match 35.3%; Score 6; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 IRKEIL 15
|||||
Db 69 IRKEIL 74

RESULT 4

H75168
2-ketoglutarate ferredoxin oxidoreductase, chain gamma (korg-2) PAB0348 - Pyrococcus aby
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: H75168
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: H75168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-185 <KAW>
A:Cross-references: GB:AJ248284; GB:AL096836; NID:95457730; PIDN:CAB49431.1; PID:9545794
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: korg-2; PAB0348
C:Superfamily: pyruvate synthase gamma chain

Query Match 35.3%; Score 6; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 RKEILI 16
|||||
Db 5 RKEILI 10

RESULT 5

A42710
superoxide dismutase (EC 1.15.1.1) (Fe/Mn) - Streptococcus mutans
C:Species: Streptococcus mutans
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 18-Jun-1999
C:Accession: A42710; A24812
R:Nakayama, K.
J. Bacteriol. 174, 4928-4934, 1992
A:Title: Nucleotide sequence of Streptococcus mutans superoxide dismutase gene and is
A:Reference number: A42710; MUID:92332426
A:Contents: GS-5
A:Accession: A42710
A:Molecule type: DNA
A:Residues: 1-203 <NAK>
A:Cross-references: GB:G39782; GB:D01037; NID:9251294; PIDN:AAB22503.1; PID:9251295
A:Note: Sequence extracted from NCBI backbone (NCBI:108940, NCBI:108941)
R:Martin, M.E.; Byers, B.R.; Olson, M.O.J.; Salin, M.L.; Arceneaux, J.E.L.; Tolbert,
J. Biol. Chem. 261, 9361-9367, 1986
A:Title: A Streptococcus mutans superoxide dismutase that is active with either manga
A:Reference number: A24812; MUID:86250886
A:Accession: A24812
A:Molecule type: protein
A:Residues: 2-3, 7, 5-23 <MAR>
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Mn)
C:Keywords: iron; manganese; metalloprotein; oxidoreductase
F:27,81,163,167/Binding site: iron/manganese (His, His, Asp, His) #status predicted

Query Match 35.3%; Score 6; DB 2; Length 203;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AILLPD 7
|||||
Db 2 AILLPD 7

RESULT 6

D64316
restriction modification enzyme subunit M1 homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: D64316
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: D64316
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-220 <BUL>
A:Cross-references: GB:U67470; GB:L77117; NID:92826247; PIDN:AAB98113.1; PID:91592267
C:Genetics:
A:Map position: REV127472-126810
A:Start codon: TTG

Query Match 35.3%; Score 6; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 KIRKEI 14
|||||
Db 66 KIRKEI 71

RESULT 7

E71653
hypothetical protein RP499 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: E71653
R:Andersson, S.G.B.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: E71653
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-247 <AND>
A:Cross-references: GB:AJ235272; GB:AJ235269; NID:g3861033; PIDN:CAA14951.1; PID:e134279
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: RP499

Query Match 35.3%; Score 6; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 IRKEIL 15
|||||
Db 80 IRKEIL 85

RESULT 8
H64395
hypothetical protein MJ0768 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: H64395
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: H64395
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-249 <BUL>
A:Cross-references: GB:U67522; GB:L77117; NID:g2826315; PIDN:AAB98763.1; PID:g1499588; T
C:Genetics:
A:Map position: FOR689591-690340
C:Superfamily: hypothetical protein MJ0768

Query Match 35.3%; Score 6; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 KIRKEI 14
|||||
Db 136 KIRKEI 141

RESULT 9
C72572
hypothetical protein APE1860 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: C72572
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai,
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
A:Reference number: A72450; MUID:99310339
A:Accession: C72572
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-251 <KAW>
A:Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80864.1; PID:d1044650; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1860
C:Superfamily: Aeropyrum pernix hypothetical protein APE1860

Query Match 35.3%; Score 6; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAILLP 6
|||||
Db 123 GAILLP 128

RESULT 10
F72264
hypothetical protein TM1350 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: F72264
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: F72264
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <ARN>
A:Cross-references: GB:AE001789; GB:AE000512; NID:g4981904; PIDN:AAD36421.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1350
C:Superfamily: peroxidase

Query Match 35.3%; Score 6; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AILLPD 7
|||||
Db 113 AILLPD 118

RESULT 11
H69874
conserved hypothetical protein ylbK - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: H69874
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portele
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033
A:Accession: H69874

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-260 <KUN>
A:Cross-references: GB:299111; GB:AL009126; NID:g2633699; PIDN:CAB13377.1; PID:el185094;
A:Experimental source: strain 168
C:Genetics:
A:Gene: ylbk

Query Match 35.3%; Score 6; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KIRKEI 14
|||||
DB 249 KIRKEI 254

RESULT 12
S78172
Probable transport protein yejU - Reclinomonas americana (ATCC 50394) mitochondrion
C:Species: Mitochondrion Reclinomonas americana
A:Variety: ATCC 50394
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Aug-1999
C:Accession: S78172
R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sank
Nature 387, 493-497, 1997
A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.
A:Reference number: S78127; MUID:97311393
A:Accession: S78172
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <LAN>
A:Cross-references: EMBL:AF007261; NID:g2258325; PIDN:AADI1905.1; PID:g2258371
A:Experimental source: ATCC 50394
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997
C:Genetics:
A:Gene: yejU
A:Gene: yejU
A:Gene: mitochondrion
C:Function:
A:Description: involved in cytochrome c1 biosynthesis
C:Superfamily: helc protein
C:Keywords: mitochondrion

Query Match 35.3%; Score 6; DB 2; Length 262;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 RKEILI 16
|||||
DB 237 RKEILI 242

RESULT 13
A69000
conserved hypothetical protein MTH1 - Methanobacterium thermoautotrophicum (strain Delta
C:Species: Methanobacterium thermoautotrophicum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A69000
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514
A:Accession: A69000
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <MTH>
A:Cross-references: GB:AE000795; GB:AE000666; NID:g2621036; PIDN:AAB84510.1; PID:g262103
A:Experimental source: strain Delta H
C:Genetics:

A:Superfamily: conserved hypothetical protein MTH1
C:Gene: MTH1

Query Match 35.3%; Score 6; DB 1; Length 268;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 KEILIE 17
|||||
DB 149 KEILIE 154

RESULT 14
C69380
hypothetical protein AF1043 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69380
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: C69380
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-269 <KLE>
A:Cross-references: GB:AE001031; GB:AE000782; NID:g2689354; PIDN:AAB90208.1; PID:g264

Query Match 35.3%; Score 6; DB 2; Length 269;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KIRKEI 14
|||||
DB 245 KIRKEI 250

RESULT 15
A41382
UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) [validated] - Acetobacter p
N:Alternate names: UDP-glucose pyrophosphorylase
C:Species: Acetobacter pasteurianus
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 31-Mar-2000
C:Accession: A41382
R:Brede, G.; Fjaervik, E.; Valla, S.
J. Bacteriol. 173, 7042-7045, 1991
A:Title: Nucleotide sequence and expression analysis of the Acetobacter xylinum uridi
A:Reference number: A41382; MUID:92041596
A:Accession: A41382
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-284 <BRE>
A:Cross-references: GB:M76548; NID:g141736; PIDN:AAA21888.1; PID:g141737
C:Function:
A:Description: catalyzes the reaction of UTP and glucose-1-phosphate to form UDP-1-gl
C:Superfamily: Escherichia coli UTP--glucose-1-phosphate uridylyltransferase
C:Keywords: nucleotidyltransferase

Query Match 35.3%; Score 6; DB 1; Length 284;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AILLPD 7
|||||
DB 129 AILLPD 134

Search completed: January 31, 2002, 13:20:02
Job time: 103 sec

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:03 ; Search time 140.03 Seconds
(without alignments)
4.232 Million cell updates/sec

Title: US-08-957-709-69

Perfect score: 8

Sequence: 1 PDWKIRKE 8

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_1101.*

1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	100.0	8	19 AAW72846	Polymerase enhanci
2	8	100.0	15	19 AAW72845	Polymerase enhanci
3	8	100.0	156	19 AAW72847	Polymerase enhanci
4	6	75.0	339	22 AAG72102	Human olfactory re
5	6	75.0	595	19 AAW76441	Human p53 regulate
6	6	75.0	595	22 AAB28243	Rat P2X ₇ /P2Z #1.
7	6	75.0	595	22 AAB28253	Rat P2X ₇ /P2Z #2.
8	5	62.5	20	22 AAB74118	Rat C5a peptide fr
9	5	62.5	20	22 AAB28242	Rat P2X ₇ /P2Z C-te
10	5	62.5	46	20 AAY25133	B. taurus CEP prot
11	5	62.5	48	22 AAM37005	Peptide #11042 enc

12	5	62.5	55	19	AAV21221	Human bcl2 proto-o
13	5	62.5	74	21	AAAG32683	ze a mays protein f
14	5	62.5	77	22	AAE05458	Mouse C5a anaphyla
15	5	62.5	77	22	AAE74051	Rat C5a. Rattus s
16	5	62.5	85	21	AAZ22515	ze a mays protein f
17	5	62.5	93	12	AAAR13468	Part of yeast proth
18	5	62.5	104	21	AAAG22514	ze a mays protein f
19	5	62.5	108	21	AAAB41320	Human ORFX ORF1084
20	5	62.5	113	21	AAAG01628	Human secreted pro
21	5	62.5	114	21	AAAG32682	ze a mays protein f
22	5	62.5	118	20	AAV35816	Amino acid sequenc
23	5	62.5	120	21	AAAG04334	Arabidopsis thalia
24	5	62.5	120	21	AAAG39817	Arabidopsis thalia
25	5	62.5	122	21	AAAG04333	Arabidopsis thalia
26	5	62.5	122	21	AAAG39816	Arabidopsis thalia
27	5	62.5	124	21	AAAG01945	Human secreted pro
28	5	62.5	134	21	AAAG04332	Arabidopsis thalia
29	5	62.5	134	21	AAAG39815	Arabidopsis thalia
30	5	62.5	136	21	AAAG35445	ze a mays protein f
31	5	62.5	148	21	AAAB56177	Human secreted pro
32	5	62.5	154	21	AAAG16696	Arabidopsis thalia
33	5	62.5	159	21	AAAB56302	Human secreted pro
34	5	62.5	183	21	AAAG35444	ze a mays protein f
35	5	62.5	191	19	AAW77765	Formate dehydrogen
36	5	62.5	199	18	AAW28199	Staphylococcus aur
37	5	62.5	206	21	AAAG35443	ze a mays protein f
38	5	62.5	208	19	AAW71749	Human ubiquitin co
39	5	62.5	208	20	AAV25129	Human UCSP-2 prote
40	5	62.5	208	21	AAV91052	Human ubiquitin co
41	5	62.5	208	22	AAAB98227	Human zinc finger
42	5	62.5	208	22	AAAB64506	Human secreted pro
43	5	62.5	210	20	AAV07028	Breast cancer asso
44	5	62.5	213	20	AAV42772	Rat neuronal immed
45	5	62.5	213	22	AAAB98228	Human zinc finger

ALIGNMENTS

RESULT 1
AAW72846
ID AAW72846 standard; Peptide; 8 AA.
XX
AC AAW72846;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 component N-terminal peptide.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
PN WO9842860-AL.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX (STRA-) STRATAGENE.
XX Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
PT Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication
XX

PS Claim 17; Page 33; 16lpp; English.

XX This peptide comprises an N-terminal peptide of the P45 component
 CC of the polymerase enhancing factor (PEF) of *Pyrococcus furiosus*
 CC DSM 3638. It has been obtained from isolated P45. PCR primers
 CC (see AAV63861-64) based on the peptide were used to amplify P45 DNA
 CC (see AAV63860). A full-length P45 sequence is provided in AAW72847.
 CC PEF, a predominant component of which is P45 protein, enhances the
 CC activity of *P. furiosus* DNA polymerase, thereby providing
 CC replication products of greater length and purity. P45 is a
 CC dUTPase and possesses polymerase enhancing activity. The invention
 CC provides novel extracts, proteins and complexes that improve the
 CC polymerisation activity of nucleic acid polymerases. These
 CC complexes may include proteins including the P45 N-terminal
 CC peptide. Also included are methods for identifying compositions
 CC with polymerase enhancing activity, for purifying and using these
 CC compositions, and specific extracts, proteins and complexes that
 CC function to enhance polymerase activity. Nucleic acid polymerase
 CC reactions can be enhanced (claimed) by mixing a nucleic acid
 CC template, at least 1 polymerase and a composition having polymerase
 CC enhancing activity. Kits are provided for replicating nucleic
 CC acids. The kits can be used in site-directed mutagenesis, nucleic
 CC acid sequencing or amplification (preferably PCR or RT-PCR).
 XX SQ Sequence 8 AA;

Query Match 100.0%; Score 8; DB 19; Length 8;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDWKIRKE 8
 |||||
 Db 1 pdwkirke 8

RESULT 2
 AAW72845
 ID AAW72845 standard; Peptide; 15 AA.

AC AAW72845;

XX 01-MAR-1999 (first entry)

XX Polymerase enhancing factor P45 component N-terminal peptide.

XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 KW amplification; sequencing; replication.

XX *Pyrococcus furiosus* strain DSM 3638.

XX Key Location/Qualifiers
 FT Misc-difference 1 /label= Gly, Ala, Ile

XX WO9842860-A1.

XX 01-OCT-1998.

XX 20-MAR-1998; 98WO-US05497.

XX 24-OCT-1997; 97US-0957709.

XX 21-MAR-1997; 97US-0822774.

XX (STRA-) STRATAGENE.

XX Hansen CJ, Hogrefe H;

XX WPI; 1998-542284/46.

XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 PT use in amplification, sequencing and replication

XX Claim 17; Page 33; 16lpp; English.

XX This peptide comprises an N-terminal peptide of the P45 component
 CC of the polymerase enhancing factor (PEF) of *Pyrococcus furiosus*
 CC DSM 3638. It has been obtained from isolated P45; a full-length
 CC sequence is provided in AAW72847. PEF, the predominant components
 CC of which are P45 and P50 (see AAW72844) proteins, enhances the
 CC activity of *P. furiosus* DNA polymerase, thereby providing
 CC replication products of greater length and purity. P45 is a
 CC dUTPase and possesses polymerase enhancing activity. The invention
 CC provides novel extracts, proteins and complexes that improve the
 CC polymerisation activity of nucleic acid polymerases. These
 CC complexes may include proteins including the P45 N-terminal
 CC peptide. Also included are methods for identifying compositions
 CC with polymerase enhancing activity, for purifying and using these
 CC compositions, and specific extracts, proteins and complexes that
 CC function to enhance polymerase activity. Nucleic acid polymerase
 CC reactions can be enhanced (claimed) by mixing a nucleic acid
 CC template, at least 1 polymerase and a composition having polymerase
 CC enhancing activity. Kits are provided for replicating nucleic
 CC acids. The kits can be used in site-directed mutagenesis, nucleic
 CC acid sequencing or amplification (preferably PCR or RT-PCR).
 XX SQ Sequence 15 AA;

Query Match 100.0%; Score 8; DB 19; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.0019;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDWKIRKE 8
 |||||
 Db 4 pdwkirke 11

RESULT 3

AAW72847
 ID AAW72847 standard; Protein; 156 AA.

AC AAW72847;

XX 01-MAR-1999 (first entry)

XX Polymerase enhancing factor P45 (dUTPase) component.

XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 KW amplification; sequencing; replication.

XX *Pyrococcus furiosus* strain DSM 3638.

XX Key Location/Qualifiers
 FT Peptide 2..15
 FT /note= "N-terminal peptide used to generate
 FT primers"

XX WO9842860-A1.

XX 01-OCT-1998.

XX 20-MAR-1998; 98WO-US05497.

XX 24-OCT-1997; 97US-0957709.

XX 21-MAR-1997; 97US-0822774.

XX (STRA-) STRATAGENE.

XX Hansen CJ, Hogrefe H;

XX WPI; 1998-542284/46.

XX N-PSDB; AAV63860.

XX Polymerase enhancing factor proteins, extracts and complexes -

PT improve the polymerisation activity of nucleic acid polymerases, for
 PT use in amplification, sequencing and replication

XX Claim 17; Page 43; 161pp; English.

XX This is the amino acid sequence of the P45 component of the
 CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.
 CC The sequence is predicted from a DNA sequence (see AAV63860) obtained
 CC from genomic DNA by PCR. P45 and P50 (see AAW72844) are the
 CC predominant components of PEF, which acts to enhance the activity of
 CC P. furiosus DNA polymerase. P45 functions as a dUTPase, and can be
 CC used to enhance nucleic acid replication, polymerisation or PCR
 CC reactions. The invention provides novel extracts, proteins and
 CC complexes that improve the polymerisation activity of nucleic acid
 CC polymerases, as well as DNA constructs and antibodies. Also
 CC included are methods for identifying compositions with PEF activity,
 CC for purifying and using these compositions, and specific extracts,
 CC proteins and complexes that function to enhance polymerase activity.
 CC Nucleic acid polymerase reactions can be enhanced (claimed) by
 CC mixing a nucleic acid template, at least 1 polymerase and a
 CC composition having polymerase enhancing activity. Kits are
 CC provided for replicating nucleic acids, for site-directed
 CC mutagenesis, for nucleic acid sequencing or amplification
 CC (preferably PCR or RT-PCR).

XX Sequence 156 AA;

Query Match 100.0%; Score 8; DB 19; Length 156;

Best Local Similarity 100.0%; Pred. No. 0.014; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDWKIRKE 8

Db 4 pdwkirke 11

RESULT 4

AAAG72102

ID AAG72102 standard; Protein; 339 AA.

XX

AC AAG72102;

XX 31-JUL-2001 (first entry)

XX Human olfactory receptor polypeptide, SEQ ID NO: 1783.

XX Human; olfactory receptor; OR; primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

KW scent profile; scent fingerprint; scent representation.

XX

OS Homo sapiens.

XX WO200127158-A2.

PN 19-APR-2001.

XX 06-OCT-2000; 2000WO-0527582.

PF 08-OCT-1999; 99US-0158615.

XX 24-FEB-2000; 2000US-0184809.

PR

XX (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

XX

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

XX WPI; 2001-290713/30.

DR

XX New polynucleotides which encode polypeptides involved in olfactory

PT sensation for identifying olfactory agonists and antagonists -

XX

PS Claim 11; Page 1172-1173; 1857pp; English.

XX

CC The present sequence is an olfactory receptor which is encoded by
 CC one of a number of novel polynucleotides. The polynucleotides can be
 CC used in screening for olfactory agonists and antagonists. The methods
 CC allow for the determination of primary scents and the identification
 CC of the odour receptors used to detect these primary scents. The methods
 CC also enable determination of secondary scents and the identification of
 CC combinations of odour receptors that are involved in detecting such
 CC secondary scents. This enables the construction of a scent representation
 CC (also called a scent fingerprint or scent profile), which may be used to
 CC re-create and edit scents. Libraries of olfactory receptors are useful
 CC for determining the interaction pattern of a composition with the
 CC receptors, and can be used for determining differences in the olfactory
 CC faculties of different individuals.

XX Sequence 339 AA;

Query Match 75.0%; Score 6; DB 22; Length 339;

Best Local Similarity 100.0%; Pred. No. 6.5;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDWKIR 6

Db 327 pdwkir 332

RESULT 5

AAW76441

ID AAW76441 standard; Protein; 595 AA.

XX

AC AAW76441;

XX 28-APR-1999 (first entry)

XX Human p53 regulated protein, P2X7.

DE

XX

KW Human; p53 regulated protein; P2XM; ATP receptor; RP-2 protein;

KW thymocyte; apoptosis; tumour-suppressor gene; rhabdoid sarcoma;

KW anticancer drug.

XX

OS Homo sapiens.

XX WO9842835-A1.

PN 01-OCT-1998.

XX

PD 18-MAR-1998; 98WO-JP01146.

XX 26-MAR-1997; 97JP-0093044.

XX (SAKA) OTSUKA PHARM CO LTD.

PA Nakamura Y, Tokino T;

XX WPI; 1998-532006/45.

XX

PI Human gene P2XM whose transcription is induced by p53 - useful, e.g.

XX for diagnostic purposes and in development of new anticancer drugs

XX

DR Disclosure; Fig 5-6; 43pp; Japanese.

XX

XX This sequence represents the human P2X7 protein. The invention relates to

XX the P2XM protein, which is significantly homologous to: (i) the P2X

XX family of ATP receptors, and (ii) RP-2 protein which is expressed in

XX thymocytes during apoptosis. Transcription of the genes is specifically

XX regulated by the tumour-suppressor gene p53. The P2XM gene is

XX specifically expressed in skeletal muscle and has been localised to

XX chromosome 22q11, an area where mutation and sequence losses frequently

XX occur in rhabdoid sarcomas. The genes may be used for diagnostic purposes

XX (e.g. by detecting changes occurring in the gene in sarcomas), using

XX probes and primers containing or derived from all or part of the genes.

XX The genes may further be used in the development of new anticancer drugs.

XX

```

XX SQ Sequence 595 AA;
Query Match 75.0%; Score 6; DB 19; Length 595;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRKE 8
Db 575 wkirke 580
|||||

RESULT 6
AAB28243
ID AAB28243 standard; Protein; 595 AA.
XX
AC AAB28243;
XX
DT 13-FEB-2001 (first entry)
DE
DE Rat P2X7/P2Z #1.
XX
KW Rat; P2X7; neuroprotective; nootropic; antiinflammatory; antirheumatic;
KW antiarthritic; antibacterial; antiviral; antiallergic; cytostatic;
KW cardiant; cerebroprotective; immunosuppressive; P2Z; purinergic receptor;
KW nervous system disorder; chronic inflammation; Alzheimer's disease;
KW rheumatoid arthritis; amyloidosis; bacterial; viral; microbial infection;
KW haematopoietic system disorder; immune response; autoimmune disorder;
KW allergy; lymphoproliferative disorder; cardiac; cerebral ischaemia;
KW tuberculosis.
XX
OS Rattus sp.
XX
PN US6133434-A.
XX
PD 17-OCT-2000.
XX
PF 28-APR-1997; 97US-0842079.
XX
PR 28-APR-1997; 97US-0842079.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Buell GN, Kawashima E, Surprenant A;
XX
DR WPI; 2001-006153/01.
XX
DR N-PSDB; AAC63693.
XX
Mammalian purinergic receptor (P2X7) useful for screening for
PT modulators which are useful for treating arthritic, respiratory
PT disorders and neurodegenerative disorders, and to generate receptors
PT specific antibodies -
XX
PS Claim 10; Fig 1B; 40pp; English.
XX
The present sequence is rat purinergic receptor P2X7/P2Z. The coding
CC sequence for this protein can be used to treat disorders of the nervous
CC system, particularly diseases with a component of chronic inflammation,
CC such as Alzheimer's disease, diseases involving acute or chronic
CC inflammation such as rheumatoid arthritis, amyloidosis, bacterial, viral
CC and other microbial infections, disorders of the haematopoietic system
CC and immune response such as autoimmune disorders, allergies and
CC lymphoproliferative disorders, diseases involving apoptotic cell death,
CC such as cardiac and cerebral ischaemia and microbial infections,
CC particularly tuberculosis.
XX
SQ Sequence 595 AA;

Query Match 75.0%; Score 6; DB 22; Length 595;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRKE 8
Db 575 wkirke 580
|||||

RESULT 7
AAB28253
ID AAB28253 standard; Protein; 595 AA.
XX
AC AAB28253;
XX
DT 13-FEB-2001 (first entry)
DE
DE Rat P2X7/P2Z #2.
XX
KW Rat; P2X7; neuroprotective; nootropic; antiinflammatory;
KW antiarthritic; antibacterial; antiviral; antiallergic; cytostatic;
KW cardiant; cerebroprotective; immunosuppressive; P2Z; purinergic receptor;
KW nervous system disorder; chronic inflammation; Alzheimer's disease;
KW rheumatoid arthritis; amyloidosis; bacterial; viral; microbial infection;
KW haematopoietic system disorder; immune response; autoimmune disorder;
KW allergy; lymphoproliferative disorder; cardiac; cerebral ischaemia;
KW tuberculosis; antirheumatic.
XX
OS Rattus sp.
XX
PN US6133434-A.
XX
PD 17-OCT-2000.
XX
PF 28-APR-1997; 97US-0842079.
XX
PR 28-APR-1997; 97US-0842079.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Buell GN, Kawashima E, Surprenant A;
XX
DR WPI; 2001-006153/01.
XX
Mammalian purinergic receptor (P2X7) useful for screening for
PT modulators which are useful for treating arthritic, respiratory
PT disorders and neurodegenerative disorders, and to generate receptors
PT specific antibodies -
XX
PS Example 3; Fig 5; 40pp; English.
XX
The present sequence is rat purinergic receptor P2X7/P2Z. The coding
CC sequence for this protein can be used to treat disorders of the nervous
CC system, particularly diseases with a component of chronic inflammation,
CC such as Alzheimer's disease, diseases involving acute or chronic
CC inflammation such as rheumatoid arthritis, amyloidosis, bacterial, viral
CC and other microbial infections, disorders of the haematopoietic system
CC and immune response such as autoimmune disorders, allergies and
CC lymphoproliferative disorders, diseases involving apoptotic cell death,
CC such as cardiac and cerebral ischaemia and microbial infections,
CC particularly tuberculosis.
XX
SQ Sequence 595 AA;

Query Match 75.0%; Score 6; DB 22; Length 595;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRKE 8
Db 575 wkirke 580
|||||

RESULT 8
AAB74118

```

ID AAB74118 standard; Peptide; 20 AA.
 AC AAB74118;
 XX
 DT 16-MAY-2001 (first entry)
 XX
 DE Rat C5a peptide fragment #2.
 XX
 XX Rat; C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX
 OS Rattus sp.
 XX
 PN WO200115731-A1.
 XX
 PD 08-MAR-2001.
 XX
 PF 31-AUG-2000; 2000WO-US24219.
 XX
 XX 31-AUG-1999; 99US-0387671.
 XX
 PA (UNWI) UNIV MICHIGAN.
 XX
 PI Ward PA, Huber-Lang M, Sarma V;
 XX
 DR WPI; 2001-226665/23.
 DR N-PSDB; RAA75799.
 XX
 XX Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides -
 XX
 PS Claim 29; Page 27; 84pp; English.
 XX
 CC The present sequence is a peptide fragment of rat complement component
 CC C5a (the full-length sequence is given in AAB74051). The present
 CC invention relates to an antibody specific for the present sequence. The
 CC C5a-antibody can be used in a therapeutic composition, which is useful
 CC for treating a subject suffering from bacterial infection, e.g.
 CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract
 CC infections or infections in wounds. In addition, the C5a antibody can
 CC be used for treating sepsis. C5a is also known as anaphylatoxin.
 XX
 SQ Sequence 20 AA;
 Query Match 62.5%; Score 5; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 KIRKE 8
 DB 6 kirke 10
 RESULT 9
 AAB28242
 ID AAB28242 standard; Peptide; 20 AA.
 AC AAB28242;
 XX
 DT 13-FEB-2001 (first entry)
 XX
 DE Rat P2X₇/P2Z C-terminal peptide.
 XX
 XX Rat; P2X₇; neuroprotective; nootropic; antiinflammatory; antirheumatic;
 KW antiarthritic; antibacterial; antiviral; antiallergic; cytostatic;
 KW cardiant; cerebroprotective; immunosuppressive; P2Z; purinergic receptor;
 KW nervous system disorder; chronic inflammation; Alzheimer's disease;
 KW rheumatoid arthritis; amyloidosis; bacterial; viral; microbial infection;
 KW hematopoietic system disorder; immune response; autoimmune disorder;
 KW allergy; lymphoproliferative disorder; cardiac; cerebral ischaemia;

KW tuberculosis.
 XX
 OS Rattus sp.
 XX
 PN US6133434-A.
 XX
 PD 17-OCT-2000.
 XX
 XX 28-APR-1997; 97US-0842079.
 PF
 XX 28-APR-1997; 97US-0842079.
 PR
 XX (GLAX) GLAXO GROUP LTD.
 PA
 XX Buell GN, Kawashima E, Surprenant A;
 PI
 XX WPI; 2001-006153/01.
 DR
 XX Mammalian purinergic receptor (P2X₇) useful for screening for
 PT modulators which are useful for treating arthritic, respiratory
 PT disorders and neurodegenerative disorders, and to generate receptors
 PT specific antibodies -
 XX
 XX Example 4; Column 15; 40pp; English.
 PS
 XX The present sequence is a C-terminal peptide of rat purinergic receptor
 CC P2X₇/P2Z. The coding sequence for P2X₇ protein (AAC63693) can be used
 CC to treat disorders of the nervous system, particularly diseases with a
 CC component of chronic inflammation, such as Alzheimer's disease, diseases
 CC involving acute or chronic inflammation such as rheumatoid arthritis,
 CC amyloidosis, bacterial, viral and other microbial infections, disorders
 CC of the hematopoietic system and immune response such as autoimmune
 CC disorders, allergies and lymphoproliferative disorders, diseases
 CC involving apoptotic cell death, such as cardiac and cerebral ischaemia
 CC and microbial infections, particularly tuberculosis. The present
 CC sequence was used to produce a rabbit polyclonal antiserum, which was
 CC used in immunohistochemistry of rat P2X₇.
 XX
 SQ Sequence 20 AA;
 Query Match 62.5%; Score 5; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 KIRKE 8
 DB 1 kirke 5
 RESULT 10
 AAY25133
 ID AAY25133 standard; Protein; 46 AA.
 XX
 XX AAY25133;
 AC
 XX
 DT 26-AUG-1999 (first entry)
 XX
 DE B. taurus CEP protein.
 XX
 XX Ubiquitin conjugation system; UGSP-1; USCP-2; pathway; eukaryotic cell;
 KW cellular protein degradation; regulation; regulatory protein; modulator;
 KW gene transcription; cell cycle progression; mitotic cyclic kinase;
 KW oncoprotein; tumour suppressor gene; viral protein; signal transduction;
 KW cell surface receptor; transcriptional regulator; treatment; cancer;
 KW astrocytoma; melanoma; lymphoma; breast; brain; prostate; angina; asthma;
 KW immune disorder; smooth muscle disorder; anaphylactic shock; measles;
 KW Cushing's syndrome; viral infection; gastroenteritis; Epstein-Barr virus;
 KW hepatitis; antigen; human; CEP protein; bovine.
 XX
 OS Bos taurus.
 XX
 PN US5922318-A.

```

XX PD 13-JUL-1999.
XX PF 13-AUG-1998; 98US-0134596.
XX PR 21-MAY-1997; 97US-0861269.
XX PR 13-AUG-1998; 98US-0134596.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Bandman O, Corley NC, Lal P, Shah P;
XX DR WPI; 1999-404435/34.
XX PT Isolated ubiquitin conjugation system protein useful for the
XX PT treatment of cancer, immune disorders smooth muscle disorders and
XX PT viral infections
XX PS Disclosure; Fig 4; 40pp; English.
XX CC This invention describes novel isolated and purified ubiquitin
XX CC conjugation system (UCS) proteins, designated UCSP-1 and UCSP-2. The
XX CC UCS proteins are part of the UCS, a major pathway for the degradation
XX CC of cellular proteins in eukaryotic cells. The UCS mediates the
XX CC elimination of abnormal proteins and regulates the half-lives of
XX CC important regulatory proteins that control cellular processes such as
XX CC gene transcription and cell cycle progression. The UCS is implicated in
XX CC the degradation of mitotic cyclic kinases, oncoproteins, tumour
XX CC suppressor genes (e.g. p53), viral proteins, cell surface receptors
XX CC associated with signal transduction, transcriptional regulators and other
XX CC mutated or damaged proteins. Therefore, UCSP-1 proteins may be used to
XX CC treat disorders associated with the inappropriate expression or activity
XX CC of UCS proteins. For example, UCSP-1 may be administered to treat cancer
XX CC (e.g. astrocytomas, melanomas, lymphomas and cancers of the breast, brain
XX CC and prostate), immune disorders, smooth muscle disorders (e.g. angina,
XX CC anaphylactic shock asthma and Cushing's syndrome) and viral infections
XX CC (e.g. measles, gastroenteritis, Epstein-Barr virus infections and
XX CC hepatitis). The UCSP-1 proteins may also be used as antigens in the
XX CC production of antibodies or in assays to identify modulators of UCS
XX CC function. This sequence represents the bovine CEP protein which is used
XX CC in the description of the invention.
XX SQ Sequence 46 AA;

Query Match 62.5%; Score 5; DB 20; Length 46;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
Db 30 Kirke 34

RESULT 11
AAM37005
ID AAM37005 standard; Protein; 48 AA.
XX AC AAM37005;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #11042 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PA

```

```

PF 30-JAN-2001; 2001WO-US00663.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 27; SEQ ID No 37274; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AA13315-AA157546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX SQ Sequence 48 AA;

Query Match 62.5%; Score 5; DB 22; Length 48;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
Db 32 Kirke 36

RESULT 12
AAY21221
ID AAY21221 standard; Protein; 55 AA.
XX AC AAY21221;
XX DT 22-JUL-1999 (first entry)
XX DE Human bcl2 proto-oncogene mutant protein fragment 69.
XX KW Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
XX KW frameshift mutation; age-related disease; neurodegenerative disorder;
XX KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
XX KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
XX KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
XX KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
XX KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
XX KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HUPF-1;
XX KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
XX KW high mobility group protein-C; neuroendocrine specific protein A.
XX OS Synthetic.
XX OS Homo sapiens.
XX PN WO9845322-A2.
XX PD 15-OCT-1998.
XX PF 02-APR-1998; 98WO-IB00705.
XX PR 10-APR-1997; 97US-0043163.
XX PA (UYUT-) RIJKSUNIV UTRECHT.

```

PA (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA (UYRO-) UNIV ROTTERDAM ERASMUS.

PI Burbach JPH, Grosveld FG, Van Leeuwen FW;

XX WPI; 1998-609901/51.

DR N-PSDB; AAX75766.

XX Diagnosing disease by detecting frameshift mutations in RNA or

PT corresponding protein mutations - used to diagnose cancer and

PT neurological diseases, particularly Alzheimer's disease, and also

PT for treatment and prevention with specific ribozymes or wild-type

PT RNA

XX Disclosure; Figure 15; 258pp; English.

XX This invention describes a novel method for the diagnosis of a disease
CC caused by, or associated with, an RNA molecule that has a frameshift
CC mutation. The method is used to diagnose age-related diseases, especially
CC cancer and a wide range of neurodegenerative disorders (e.g. Alzheimer's
CC disease, Down's syndrome, myotonic dystrophy, Huntington's disease,
CC multiple sclerosis, alcoholic liver disease, diabetes mellitus type II
CC and many others listed) or susceptibility to these disorders. The method
CC allows a definitive diagnosis of Alzheimer's disease in living patients,
CC at an early stage. It is based on the observation that disease may be
CC caused by mutations in RNA rather than DNA. The invention describes the
CC use of neuronal system RNA molecules, specifically proteins including
CC beta-amyloid precursor protein (beta-APP), the microtubule associated
CC proteins Tau and Big Tau, ubiquitin B, apolipoprotein E, microtubule
CC associated protein 2 (MAP2), neurofilament-L, neurofilament-M,
CC neurofilament-F, presenilin I, presenilin II, glial fibrillary acidic
CC protein (GFAP), the cellular tumour antigen p53, B-cell leukemia/lymphoma
CC 2 (bcl-2) proto-oncogene, semaphorin III, HUPF-1, high mobility group
CC protein-C (HMGP-C) and neuroendocrine specific protein A.

XX Sequence 55 AA;

Query Match 62.5%; Score 5; DB 19; Length 55;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KIRKE 8

Db 30 Kirke 34

RESULT 13

AAG32683
ID AAG32683 standard; Protein; 74 AA.

XX AC AAG32683;

XX DT 17-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 39475.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.

XX Zea mays subsp. mays.

OS EP1033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

```

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151090.
PR 30-AUG-1999; 99US-0151393.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152353.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153358.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.5%; Score 5; DB 21; Length 74;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 4 KIRKE 8
Db 40 kirke 44
|||||

RESULT 14
AAE05458
ID AAE05458 standard; Protein; 77 AA.
XX AC AAE05458;
XX DT 24-SEP-2001 (first entry)
XX DE Mouse C5a anaphylatoxin.
XX KW Mannan binding protein-associated serine protease; MASP;
XX KW complement-activation; C-activation; microbial infection;
XX KW cytomegalovirus; CMV; hepatitis virus; human immunodeficiency virus;
XX KW HIV; organ transplant rejection; tissue injury; autoimmune disease;
XX KW rheumatoid arthritis; systemic lupus erythematosus; SLE; mouse;
XX KW inflammatory response; Alzheimer's disease; C5a anaphylatoxin.
XX OS Mus sp.
XX PN US6235494-B1.
XX PD 22-MAY-2001.
XX PF 08-FEB-1999; 99US-0246500.
XX PF 08-FEB-1999; 99US-0246500.
XX PA (SCRI ) SCRIPPS RES INST.
XX PI Hugli TE;
XX DR WPI; 2001-450082/48.
XX

```

PT Method for determining in-vivo levels of activated mannan binding
 PT protein-associated serine protease enzyme, and new peptide derivatives
 XX
 XX
 XX

PS Claim 3; Column 51-52; 39pp; English.
 XX

CC The invention relates to assays for measuring in-vivo levels of
 CC activated mannan-binding protein-associated serine protease (MASP-1 and
 CC MASP-2) activity. The assay comprises contacting a sample of blood or
 CC plasma comprising a metal ion chelator with a substrate of formula:
 CC R-peptide-Y; wherein the peptide comprises at least 4 residues from the
 CC C-terminus of an anaphylatoxin selected from C3a, C4a and C5a; R is a
 CC peptide blocking, protecting or capping group; and Y is any group
 CC cleavable from the substrate by MASP and comprises a labelled tag. The
 CC substrate of the invention is also useful for monitoring in vitro and
 CC in vivo complement-activation (C-activation) by classical, alternative
 CC or lectin pathways. The assays of MASP activity can be used for
 CC detecting or monitoring a condition associated with complement
 CC activation. The conditions include microbial infections
 CC caused by particularly cytomegalovirus (CMV), hepatitis virus and
 CC human immunodeficiency virus (HIV), organ transplant rejection, tissue
 CC injury, autoimmune diseases (e.g. rheumatoid arthritis, systemic lupus
 CC erythematosus (SLE)) and inflammatory responses (e.g. in Alzheimer's
 CC disease and bacterial diseases). The assay of MASP is also useful for
 CC assessing the toxicity or injury of therapeutic treatments or screening
 CC test compounds as agents for treatment of viral diseases, parasitic
 CC infections, tissue injury, organ transplant rejection, autoimmune
 CC diseases or inflammatory responses. The present sequence is mouse
 CC C5a anaphylatoxin.
 XX

SQ Sequence 77 AA;

Query Match 62.5%; Score 5; DB 22; Length 77;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
 Db 63 kirke 67

RESULT 15

AAAB74051
 ID AAB74051 standard; protein; 77 AA.

AC AAB74051;

DT 16-MAY-2001 (first entry)

DE Rat C5a.

KW Rat; C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.

OS Rattus sp.

PN WO200115731-A1.

XX 08-MAR-2001.

PF 31-AUG-2000; 2000WO-US24219.

XX 31-AUG-1999; 99US-0387671.

XX (UNMI) UNIV MICHIGAN.

PA Ward PA, Huber-Lang M, Sarma V;

PI WPI; 2001-226665/23.

XX DR N-PSDB; AAF75794.

XX

PT Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides -
 XX
 XX

PS Example 1; Page 26; 84pp; English.
 XX

CC The present sequence is rat complement component C5a. The present
 CC invention relates to an antibody specific for the present sequence. The
 CC C5a-antibody can be used in a therapeutic composition, which is useful
 CC for treating a subject suffering from bacterial infection, e.g.
 CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract
 CC infections or infections in wounds. In addition, the C5a antibody can
 CC be used for treating sepsis. C5a is also known as anaphylatoxin.

XX Sequence 77 AA;

Query Match 62.5%; Score 5; DB 22; Length 77;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
 Db 63 kirke 67

Search completed: January 31, 2002, 13:18:04
 Job time: 165 sec

This Page Blank (uspto)

]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 31, 2002, 13:20:05 ; Search time 78.64 seconds
(without alignments)
7.749 Million cell updates/sec

Title: US-08-957-709-69

Perfect score: 8

Sequence: 1 PDWKIRKE 8

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	100.0	154	2	dCTP deaminase (EC
2	8	100.0	156	2	dCTP deaminase (EC
3	6	75.0	478	2	leukotoxin secreti
4	6	75.0	1426	2	hypothetical prote
5	6	75.0	1972	2	TOG protein - huma
6	5	62.5	46	2	ubiquitin homolog
7	5	62.5	144	2	hypothetical prote
8	5	62.5	149	2	molybdopterin bios
9	5	62.5	167	2	NADH dehydrogenase
10	5	62.5	192	2	hypothetical prote
11	5	62.5	195	2	hypothetical prote
12	5	62.5	196	1	probable O-acetyl
13	5	62.5	210	2	DNA-directed RNA p
14	5	62.5	219	2	hypothetical prote
15	5	62.5	220	2	restriction modifi
16	5	62.5	230	2	probable 1-acylgly
17	5	62.5	233	1	orotate phosphorib
18	5	62.5	241	1	hypothetical prote
19	5	62.5	241	2	hypothetical prote
20	5	62.5	243	2	hypothetical prote
21	5	62.5	249	1	hypothetical prote
22	5	62.5	251	2	3-isopropylmalate
23	5	62.5	257	2	transcription regu
24	5	62.5	260	2	conserved hypothet
25	5	62.5	269	2	hypothetical prote
26	5	62.5	287	2	inhibitor PHB1 -
27	5	62.5	305	2	hemolysin E - Esch
28	5	62.5	305	2	probable pore form
29	5	62.5	306	2	conserved hypothet

prfD protein - Esc
plastoquinol--plas
hypothetical prote
conserved hypothet
annexin II type 1
tryptophanyl tRNA
tryptophan--tRNA 1
probable disulfide
disulfide bond iso
probable rRNA (ade
3-isopropylmalate
hypothetical prote
3-isopropylmalate
3-isopropylmalate
3-isopropylmalate
3-isopropylmalate

ALIGNMENTS

RESULT 1

G75030

dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75030

R:anonymous, Genoscope

Submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: G75030

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-154 <KAW>

A:Cross-references: GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB50685.1; PID:G545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: dcd; PAB1164

C:Superfamily: dCTP deaminase

C:Keywords: hydrolase

Query Match 100.0%; Score 8; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDWKIRKE 8

Db 4 PDWKIRKE 11

RESULT 2

E71216

dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: E71216

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: E71216

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-156 <KAW>

A:Cross-references: GB:AP000007; NID:G3236134; PIDN:BAA31124.1; PID:G3258441

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH1997

C:Superfamily: dCTP deaminase

C:Keywords: hydrolase

Query Match 100.0%; Score 8; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PDWKIRKE 8
|||||
Db 4 PDWKIRKE 11

RESULT 3
D30169
leukotoxin secretion protein lktD - Pasteurella haemolytica
C:Species: Pasteurella haemolytica
C:Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 16-Jul-1999
C:Accession: D30169; S29518
R:Highlander, S.K.; Chidambaram, M.; Engler, M.J.; Weinstock, G.M.
DNA 8, 15-28, 1989
A:Title: DNA sequence of the Pasteurella haemolytica leukotoxin gene cluster.
A:Reference number: A30169; MUID:89210283
A:Accession: D30169
A:Status: Preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-478 <Hic>
R:Strathdee, C.A.; Lo, R.Y.C.
J. Bacteriol. 171, 916-928, 1989
A:Title: Cloning, nucleotide sequence, and characterization of genes encoding the secreted leukotoxin of Pasteurella haemolytica
A:Reference number: A32051; MUID:89123172
A:Accession: B32051
A:Status: Preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-17, 'T', 19-46, 'D', 48-235, 'F', 237-265, 'EL', 268-478 <STR>
R:Lo, R.Y.C.; Strathdee, C.A.; Shewen, P.E.
Infect. Immun. 55, 1987-1996, 1987
A:Title: Nucleotide sequence of the leukotoxin genes of Pasteurella haemolytica A1.
A:Reference number: S29515; MUID:87306837
A:Accession: S29518
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-17, 'T', 19-46, 'D', 48-235, 'F', 237-265, 'EL', 268-478 <LOR>
A:Cross-references: EMBL:M20730; NID:g150492; PIDN:AAA25531.1; PID:g150496
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1988
C:Genetics:
A:Gene: lktD
C:Superfamily: hemolysin secretion protein D; lipoyl/biotin-binding homology
F:92-135, 329-358/Domain: lipoyl/biotin-binding homology #status atypical <LPB>

Query Match 75.0%; Score 6; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 6.1; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 0

Qy 3 WKIRKE 8
|||||
Db 23 WKIRKE 28

RESULT 4
T00337
hypothetical protein KIAA0568 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00337
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete coding sequence of the cDNA for KIAA0568, a novel human gene.
A:Reference number: Z14086; MUID:98290545
A:Accession: T00337
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1426 <NAG>

A:Cross-references: EMBL:AB011140; NID:g3043659; PIDN:BAA25494.1; PID:g3043660
A:Experimental source: brain; clone HH280
C:Genetics:
A:Note: KIAA0568

Query Match 75.0%; Score 6; DB 2; Length 1426;
Best Local Similarity 100.0%; Pred. No. 14; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 0

Qy 3 WKIRKE 8
|||||
Db 563 WKIRKE 568

RESULT 5
S68176
TOG protein - human
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: S68176
R:Charrasse, S.; Hazel, M.; Taviaux, S.; Berta, P.; Chow, T.; Larroque, C.
Eur. J. Biochem. 234, 406-413, 1995
A:Title: Characterization of the cDNA and pattern of expression of a new gene over-expressed in human placenta
A:Reference number: S68176; MUID:96128167
A:Accession: S68176
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1972 <CHA>
A:Cross-references: EMBL:X92474; NID:g1045056; PIDN:CAA63212.1; PID:g1045057

Query Match 75.0%; Score 6; DB 2; Length 1972;
Best Local Similarity 100.0%; Pred. No. 18; Mismatches 0; Indels 0; Gaps 0;
Matches 6; Conservative 0

Qy 3 WKIRKE 8
|||||
Db 874 WKIRKE 879

RESULT 6
I47035
ubiquitin homolog - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 05-Nov-1999
C:Accession: I47035
R:Shima, D.T.; Saunders, K.B.; Gougos, A.; D'Amore, P.A.
Differentiation 58, 217-226, 1995
A:Title: Alterations in gene expression associated with changes in the state of endototoxemia in sheep
A:Reference number: I47034; MUID:95228954
A:Accession: I47035
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-46 <SHI>
A:Cross-references: GB:S77733; NID:g998679; PIDN:AAB34029.1; PID:g998680

Query Match 62.5%; Score 5; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 14; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0

Qy 4 KIRKE 8
|||||
Db 30 KIRKE 34

RESULT 7
F69301
hypothetical protein AF0414 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Nov-1999
C:Accession: F69301

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343
A:Accession: F69301
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-144 <LE>
A:Cross-references: GB:AE001076; GB:AE000782; MID:g2689399; PIDN:AAB90831.1; PID:g265022
C:Superfamily: Archaeoglobus fulgidus hypothetical protein AF0414

Query Match 62.5%; Score 5; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 KIRKE 8
Db 62 KIRKE 66
|||||

RESULT 8
F71240
molybdopter in biosynthesis protein E chain [similarity] - Pyrococcus horikoshii
N:Alternate names: moaE protein; molybdopter in-converting factor 16K chain; molybdopter
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 15-Sep-2000
C:Accession: F71240
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137
A:Accession: F71240
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-149 <KAW>
A:Cross-references: GB:AP000001; MID:g3236128; PIDN:BAA29253.1; PID:g3256570
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0184
C:Superfamily: molybdopter in biosynthesis protein E chain
C:Keywords: molybdopter in biosynthesis

Query Match 62.5%; Score 5; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 KIRKE 8
Db 66 KIRKE 70
|||||

RESULT 9
H81252
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain B Cj1578c [imported] - Campylobacte
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 15-Sep-2000
C:Accession: H81252
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barre
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912
A:Accession: H81252
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-167 <PAR>
A:Cross-references: GB:AL139079; GB:AL111168; MID:g6968971; PIDN:CAB73566.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: nuoB; Cj1578c
C:Superfamily: psbG protein
C:Keywords: NAD; oxidoreductase

Query Match 62.5%; Score 5; DB 2; Length 167;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 KIRKE 8
Db 151 KIRKE 155
|||||

RESULT 10
B75144
hypothetical protein PAB0211 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: B75144
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: B75144
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <RAW>
A:Cross-references: GB:AJ248284; GB:AL096836; MID:g5457730; PIDN:CAB49233.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0211
C:Superfamily: conserved hypothetical protein yhcV; CBS homology
F;77-124/Domain: CBS homology <CBS>

Query Match 62.5%; Score 5; DB 2; Length 192;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 KIRKE 8
Db 15 KIRKE 19
|||||

RESULT 11
A75171
hypothetical protein PAB0359 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: A75171
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: A75171
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-195 <RAW>
A:Cross-references: GB:AJ248284; GB:AL096836; MID:g5457730; PIDN:CAB49448.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0359
C:Superfamily: hypothetical protein MJ0882

Query Match 62.5%; Score 5; DB 2; Length 195;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDWKI 5
Db 56 PDWKI 60

RESULT 12

S50709
probable O-acetyltransferase (EC 2.3.1.-) YJL218w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein HRA196; hypothetical protein J0224
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S50709; S57008; S45154
R:Vandenbol, M.; Durand, P.; Bolle, P.A.; Dion, C.; Portetelle, D.; Hilger, F.
Yeast 10, 1657-1662, 1994
A:Title: Sequence analysis of a 40.2 kb DNA fragment located near the left telomere of
A:Reference number: S50701; MUID:95242842
A:Accession: S50709
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-196 <VAN>
A:Cross-references: EMBL:Z34098; NID:q496934; PIDN:CAA83992.1; PID:q496943
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the Protein, Sequence Database, September 1995
A:Reference number: S56835
A:Accession: S57008
A:Molecule type: DNA
A:Residues: 1-196 <VAN>
A:Cross-references: EMBL:Z49493; NID:gl015607; PIDN:CAA89515.1; PID:gl015608; GSPDB:GN00
C:Genetics:
A:Map position: 10L
A:Gene: MIPS:YJL218w
A:Superfamily: galactoside acetyltransferase
C:Keywords: acyltransferase

Query Match 62.5%; Score 5; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
Db 23 KIRKE 27

RESULT 13

A32618
DNA-directed RNA polymerase (EC 2.7.7.6) II 23K chain [validated] - human
N:Alternate names: RNA polymerase II chain RPB5
C:Species: Homo sapiens (man)
C:Date: 21-May-1990 #sequence_revision 17-May-1996 #text_change 28-Jul-2000
C:Accession: S52002; A32618; A39879
R:Cheong, J.; Yi, M.; Lin, Y.; Murakami, S.
EMBO J. 14, 143-150, 1995
A:Title: Human RPB5, a subunit shared by eukaryotic nuclear RNA polymerases, binds human
A:Reference number: S52002; MUID:95129541
A:Accession: S52002
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-210 <CHE>
A:Cross-references: GB:D38251; NID:g806518; PIDN:BAA07406.1; PID:gl060912
R:Pati, U.K.; Weissman, S.M.
J. Biol. Chem. 264, 13114-13121, 1989
A:Title: Isolation and molecular characterization of a cDNA encoding the 23-kDa subunit
A:Reference number: A32618; MUID:89327280
A:Accession: A32618
A:Molecule type: mRNA
A:Residues: 1-156, 'T', 158-162, 'DKPPREPAADPGGPGCALLNDKANAGGEDHPAQ' <PAT>
A:Cross-references: GB:J04965
A:Note: this sequence has been corrected
R:Pati, U.K.; Weissman, S.M.
J. Biol. Chem. 266, 13468, 1991

A:Reference number: A39879; MUID:91302386

A:Contents: erratum

A:Accession: A39879

A:Molecule type: mRNA

A:Residues: 163-185, 'K', 187-210 <PA2>

A:Cross-references: GB:S42643; NID:g232953; PIDN:AAB19339.1; PID:g232954

C:Function:

A:Description: EC 2.7.7.6 [validated; MUID:95129541]

C:Superfamily: DNA-directed RNA polymerase chain RPB5

C:Keywords: nucleotidyltransferase; nucleus; transcription

Query Match 62.5%; Score 5; DB 2; Length 210;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRK 7
Db 11 WKIRK 15

RESULT 14

F69435
hypothetical protein AF1487 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: F69435
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
Godek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: F69435
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-219 <KLE>
A:Cross-references: GB:AE001000; GB:AE000782; NID:g2689323; PIDN:AAB89768.1; PID:g264

Query Match 62.5%; Score 5; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
Db 112 KIRKE 116

RESULT 15

D64316
restriction modification enzyme subunit M1 homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: D64316
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannas
A:Reference number: A64300; MUID:96337999
A:Accession: D64316
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-220 <BUL>
A:Cross-references: GB:U67470; GB:L77117; NID:g2826247; PIDN:AAB98113.1; PID:gl592267
C:Genetics:
A:Map position: REV127472-126810
A:Start codon: TTC

Query Match 62.5%; Score 5; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
|
|
|
|
|
Db 66 KIRKE 70

Search completed: January 31, 2002, 13:20:06
Job time: 107 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:12 ; Search time 46.78 Seconds
(without alignments)
6.270 Million cell updates/sec

Title: US-08-957-709-69
Perfect score: 8
Sequence: 1 PDWKIRKE 8

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	100.0	154	1	Q9UXS8 Pyrococcus
2	8	100.0	156	1	O57706 Pyrococcus
3	6	75.0	478	1	PL5534 Pasteurella
4	6	75.0	478	1	P55125 Pasteurella
5	6	75.0	595	1	O64663 Rattus norv
6	6	75.0	1756	1	O60437 Homo sapien
7	6	75.0	2032	1	CTOG_HUMAN
8	5	62.5	58	1	VPU_HV12H
9	5	62.5	144	1	Y414_ARCFU
10	5	62.5	196	1	YJVB_YEAST
11	5	62.5	210	1	RPB5_HUMAN
12	5	62.5	220	1	Y132_METJA
13	5	62.5	233	1	PYRE_COLGR
14	5	62.5	234	1	PYRE_METAN
15	5	62.5	241	1	YC34_METJA
16	5	62.5	249	1	Y768_METJA
17	5	62.5	260	1	YLBK_BACSU
18	5	62.5	287	1	PHB_YEAST
19	5	62.5	303	1	HLVE_ECOLI
20	5	62.5	313	1	PRRD_ECOLI
21	5	62.5	320	1	CYF_VICFA
22	5	62.5	336	1	Y883_METJA
23	5	62.5	339	1	ANXB_XENLA
24	5	62.5	344	1	SYW_CHLPN
25	5	62.5	359	1	LEU3_KLUMA
26	5	62.5	362	1	LEU3_KLULA
27	5	62.5	362	1	LEU3_PICAN
28	5	62.5	363	1	LEU3_PICJA
29	5	62.5	364	1	LEU3_YEAST
30	5	62.5	365	1	LEU3_CANBO
31	5	62.5	365	1	LEU3_CANGA
32	5	62.5	368	1	LEU3_YAMOH
33	5	62.5	370	1	YR51_CABEL

34 5 62.5 373 1 LEU3_CANAL P87186 candida alb
35 5 62.5 373 1 LEU3_CANNA P07139 candida mal
36 5 62.5 373 1 LEU3_PICST O94114 pichia stip
37 5 62.5 400 1 HEMI_CLOPE Q92nd3 clostridium
38 5 62.5 410 1 PRS1_METH O26824 methanobact
39 5 62.5 449 1 AROA_PSES2 P56952 pseudomonas
40 5 62.5 477 1 HLYD_ACTAC P18790 actinobacil
41 5 62.5 477 1 RY3D_ACTPL Q08633 actinobacil
42 5 62.5 478 1 HLY4_ECOLI P09986 escherichia
43 5 62.5 478 1 HLYD_ECOLI P07739 escherichia
44 5 62.5 512 1 GAG_SMSAV P03330 simian sarc
45 5 62.5 526 1 ALO_YEAST P54783 saccharomyc

ALIGNMENTS

RESULT 1
ID DCD_PYRAB STANDARD; PRT; 154 AA.
AC Q9UXS8: 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE).
GN DCD OR PAB1164.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ248288; CAB50685.1; -
DR InterPro: IPR003232; dCTP.deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP.deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 154 AA; 17758 MW; C031BEE419094DDB CRC64;

Query Match 100.0%; Score 8; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PDWKIRKE 8
| | | | | | | |
Db 4 PDWKIRKE 11

RESULT 2

ID DCD_PYRHO STANDARD; PRT; 156 AA.
AC O57706;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR PH1997.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka T., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuwa H., Kikuchi H.
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3";
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AP000007; BAA31124.1; -;
DR InterPro; IPR003232; dCTP_deaminase.
DR Pfam; PF00692; dUTPase.
DR ProDom; PD004900; dCTP_deaminase; 1.
DR HydroLase; Complete proteome.
SQ SEQUENCE 156 AA; 17871 MW; 96B2C2C50393A985 CRC64;

Query Match 100.0%; Score 8; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 0.0055;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PDWKIRKE 8
DB 4 PDWKIRKE 11
|||||||

RESULT 3
ID HLYD_PASHA STANDARD; PRT; 478 AA.
AC P16534;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE LEUKOTOXIN SECRETION PROTEIN D.
GN LKTD.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Mannheimia.
OX NCBI_TaxID=75985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE A1;
RX MEDLINE=87306837; PubMed=3040588;
RA Lo R.Y.C., Strathdee C.A., Shewen P.E.;
RT "Nucleotide sequence of the leukotoxin genes of Pasteurella
RT haemolytica A1";
RL Infect. Immun. 55:1987-1996(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE A1;
RX MEDLINE=89123172; PubMed=2914876;

RA Strathdee C.A., Lo R.Y.C.;
RT "Cloning, nucleotide sequence, and characterization of genes encoding
RT the secretion function of the Pasteurella haemolytica leukotoxin
RT determinant";
RL J. Bacteriol. 171:916-928(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE A1 / PHL101;
RX MEDLINE=89210283; PubMed=2707120;
RA Highlander S.K., Chidambaram M., Engler M.J., Weinstein G.M.;
RT "DNA sequence of the Pasteurella haemolytica leukotoxin gene
RT cluster";
RL DNA 8:15-28(1989).
CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF THE LEUKOTOXIN.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-BOUND (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M20730; AAA25531.1; -;
DR EMBL; M24197; AAA25545.1; -;
DR PIR; S29518; S29518.
DR InterPro; IPR002215; HlyD.
DR Pfam; PF00529; HlyD; 1.
DR PROSITE; PS00543; HLYD_FAMILY; 1.
KW Hemolysis; Cytolysis; Transport; Transmembrane; Inner membrane.
FT DOMAIN 1 59 POTENTIAL.
FT TRANSMEM 60 80
FT DOMAIN 81 478 PERIPLASMIC (POTENTIAL).
FT DOMAIN 18 18 T -> I (IN REF. 3).
FT CONFLICT 47 47 D -> E (IN REF. 3).
FT CONFLICT 150 150 N -> T (IN REF. 3).
FT CONFLICT 236 236 F -> L (IN REF. 3).
FT CONFLICT 266 267 EL -> AV (IN REF. 3).
SQ SEQUENCE 478 AA; 54761 MW; E2B8B5FF5AF988C1 CRC64;

Query Match 75.0%; Score 6; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 WKIRKE 8
DB 23 WKIRKE 28
|||||||

RESULT 4
ID HLYD_PASSP STANDARD; PRT; 478 AA.
AC P55125;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE LEUKOTOXIN SECRETION PROTEIN D.
GN LKTD.
OS Pasteurella haemolytica-like sp. (strain 5943B).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=28165;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93239320; PubMed=8478098;
RA Chang Y.-F., Ma D.-P., Shi J., Chengappa M.M.;
RT "Molecular characterization of a leukotoxin gene from a Pasteurella
RT haemolytica-like organism, encoding a new member of the RTX toxin
RT family";


```

RL Infect. Immun. 61:2089-2095(1993).
CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF THE LEUKOTOXIN.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-BOUND (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE HLYD FAMILY OF SECRETION PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L12148; AAA16446.1; -.
DR InterPro; IPR002215; HLYD.
DR Pfam; PF00529; HLYD; 1.
DR PROSITE; PS00543; HLYD_FAMILY; 1.
KW Hemolysis; Cytolysis; Transport; Transmembrane; Inner membrane.
FT DOMAIN 1 77 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 78 98 POTENTIAL.
FT DOMAIN 99 478 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 478 AA; 54650 MW; 221461A69482913A CRC64;

Query Match 75.0%; Score 6; DB 1; Length 478;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRKE 8
DB 23 WKIRKE 28

RESULT 5
P2X7_RAT
ID P2X7_RAT STANDARD; PRT; 595 AA.
AC Q84663;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE P2X PURINORECEPTOR 7 (ATP RECEPTOR) (P2X7) (PURINERGIC RECEPTOR)
DE (P2X RECEPTOR).
DE P2X7.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=M33;
RX MEDLINE=96202420; PubMed=8614837;
RA Surprenant A., Rasmussen F., Kawashima E., North R.A., Buell G.N.;
RT "The cytosolic P2X receptor for extracellular ATP identified as a P2X
RT receptor (P2X7).";
RL Science 272:735-738(1996).
CC -1- FUNCTION: RECEPTOR FOR ATP THAT ACTS AS A LIGAND GATED ION
CC CHANNEL. RESPONSIBLE FOR ATP-DEPENDENT LYSIS OF MACROPHAGES
CC THROUGH THE FORMATION OF MEMBRANE PORES PERMEABLE TO LARGE
CC MOLECULES. COULD FUNCTION IN BOTH FAST SYNAPTIC TRANSMISSION AND
CC THE ATP-MEDIATED LYSIS OF ANTIGEN-PRESENTING CELLS.
CC -1- SUBUNIT: HOMO- OR HETEROOLYMERS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE P2X RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X95882; CAA65131.1; -.

```

```

DR InterPro; IPR001429; P2X_receptor.
DR Pfam; PF00864; P2X_receptor; 1.
DR PRINTS; PR01307; P2XRECEPTOR.
DR PRINTS; PR01314; P2X7RECEPTOR.
DR PROSITE; PS01212; P2X_RECEPTOR; 1.
KW Ionic channel; Transmembrane; Ion transport; Receptor; Glycoprotein.
FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 26 46 1 (POTENTIAL).
FT DOMAIN 47 334 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 335 355 2 (POTENTIAL).
FT DOMAIN 356 595 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 213 213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 595 AA; 68391 MW; 4A6DD6058E5988D3 CRC64;

Query Match 75.0%; Score 6; DB 1; Length 595;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRKE 8
DB 575 WKIRKE 580

RESULT 6
PEPL_HUMAN
ID PEPL_HUMAN STANDARD; PRT; 1756 AA.
AC O60437; O60454; O60314;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PERIPLAKIN (195 KDA CORNIFIED ENVELOPE PRECURSOR) (190 KDA
DE PARANEOPlastic MEMPHIGUS ANTIGEN).
DE PPL OR KIAA0568.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Keratinocytes;
RX MEDLINE=98075094; PubMed=9412476;
RA Ruhrberg C., Hajibagheri M.A.N., Parry D.A.D., Watt F.M.;
RT "Periplakin, a novel component of cornified envelopes and desmosomes
RT that belongs to the plakins family and forms complexes with
RT envoplakin".
RL J. Cell Biol. 139:1835-1849(1997).
RN [2]
RN SEQUENCE FROM N.A.
RX TISSUE=Keratinocytes;
RX MEDLINE=98190524; PubMed=9521878;
RA Aho S., McLean W.H.I., Li K., Uitto J.;
RT "cDNA cloning, mRNA expression, and chromosomal mapping of human and
RT mouse periplakin genes.";
RL Genomics 48:242-247(1998).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE=99162402; PubMed=10051401;
RA Aho S., Rothenberger K., Tan E.M.L., Ryoo Y.W., Cho B.H.;
RT "Human periplakin: genomic organization in a clonally unstable region
RT of chromosome 16p with an abundance of repetitive sequence elements.";
RL Genomics 56:160-168(1999).
RN [4]
RN SEQUENCE OF 331-1756 FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,

```

RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
CC -1- FUNCTION: COMPONENT OF THE CORNIFIED ENVELOPE OF KERATINOCYTES.
CC MAY LINK THE CORNIFIED ENVELOPE TO DESMOSOMES AND INTERMEDIATE
CC FILAMENTS.
CC -1- SUBUNIT: HOMODIMER OR A HETERODIMER WITH EVPL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH DESMOSOMES AND INTERMEDIATE
CC FILAMENTS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN STRATIFIED SQUAMOUS EPITHELIA AND
CC IN SOME OTHER EPITHELIA.
CC -1- INDUCTION: DURING DIFFERENTIATION OF EPIDERMAL KERATINOCYTES.
CC -1- SIMILARITY: CONTAINS 2 PLECTIN REPEATS.
CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -1- SIMILARITY: BELONGS TO THE PLECTIN OR CYTOLINKER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF001691; AAC17738.1; -
CC EMBL: AF013717; AAC39668.1; -
CC EMBL: AF011004; AAD17459.1; -
CC EMBL: AF040999; AAD17459.1; JOINED.
CC EMBL: AF041000; AAD17459.1; JOINED.
CC EMBL: AF041002; AAD17459.1; JOINED.
CC EMBL: AF041003; AAD17459.1; JOINED.
CC EMBL: AB011140; BAA25494.1; -
CC MIN; 602871; -
CC InterPro: IPR001101; Plectin_repeat.
CC InterPro: IPR00250; PLEC; 2.
CC SMART: SM00150; SPEC; 4.
CC Repeat: Coiled coil; Cytoskeleton; Structural protein.
CC DOMAIN 16 125
CC COILED COIL (POTENTIAL).
FT DOMAIN 188 389
FT COILED COIL (POTENTIAL).
FT REPEAT 216 317
FT SPECTRIN 1.
FT REPEAT 323 485
FT SPECTRIN 2.
FT REPEAT 505 612
FT SPECTRIN 3.
FT REPEAT 733 861
FT SPECTRIN 4.
FT DOMAIN 785 820
FT COILED COIL (POTENTIAL).
FT DOMAIN 886 1645
FT COILED COIL (POTENTIAL).
FT REPEAT 1651 1685
FT PLECTIN 1.
FT REPEAT 1700 1735
FT PLECTIN 2.
FT CONFLICT 589 589
FT Q -> R (IN REF. 2 AND 3).
FT CONFLICT 657 657
FT A -> P (IN REF. 2 AND 3).
FT CONFLICT 994 994
FT V -> F (IN REF. 1).
FT CONFLICT 1573 1573
FT Q -> E (IN REF. 4).
FT CONFLICT 1663 1663
FT P -> L (IN REF. 1).
SQ SEQUENCE 1756 AA; 204649 MW; 557C7D92BE18F107 CRC64;

Query Match 75.0%; Score 6; DB 1; Length 1756;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRKE 8
Db 893 WKIRKE 898
|||||

RESULT 7
CTOG HUMAN STANDARD; PRT; 2032 AA.
AC Q14008; Q14668; PRT;
DI 15-JUL-1998 (Rel. 36, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CH-TOG PROTEIN (COLONIC AND HEPATIC TUMOR OVER-EXPRESSED PROTEIN).
GN KIAA0097.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain tumor;
RX MEDLINE=96178167; PubMed=8536682;
RA Charrasse S., Mazel M., Taviaux S., Berta P., Chow T., Larroque C.;
RT "Characterization of the cDNA and pattern of expression of a new gene
RL over-expressed in human hepatomas and colonic tumors.";
RN Eur. J. Biochem. 234:406-413(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain marrow;
RX MEDLINE=95308325; PubMed=7788527;
RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawarabayashi Y., Kotani H., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. III.
FT The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
FT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:37-43(1995).
CC -1- TISSUE SPECIFICITY: OVER-EXPRESSED IN HEPATOMAS AND COLONIC
CC TUMORS. ALSO EXPRESSED IN SKELETAL MUSCLE, BRAIN, HEART, PLACENTA,
CC LUNG, LIVER, KIDNEY AND PANCREAS.
CC -1- SIMILARITY: CONTAINS 9 HEAT REPEATS.
CC -1- SIMILARITY: TO C.ELEGANS F22B5.7 AND F22B5.8 AND SOME, TO YEAST
CC STU2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X92474; CAA63212.1; -
CC EMBL: D43948; BAA07892.2; ALT_INIT.
CC InterPro: IPR000357; HEAT_repeat.
CC PROSITE: PS00077; HEAT_REPEAT; 1.
CC Repeat.
FT REPEAT 160 197
FT HEAT 1.
FT REPEAT 357 394
FT HEAT 2.
FT REPEAT 435 472
FT HEAT 3.
FT REPEAT 751 788
FT HEAT 4.
FT REPEAT 937 974
FT HEAT 5.
FT REPEAT 1014 1051
FT HEAT 6.
FT REPEAT 1285 1322
FT HEAT 7.
FT REPEAT 1324 1357
FT HEAT 8.
FT REPEAT 1362 1399
FT HEAT 9.
FT CONFLICT 1564 1623
FT MISSING (IN REF. 1).
FT CONFLICT 1814 1814
FT E -> A (IN REF. 1).
FT CONFLICT 1822 1822
FT E -> A (IN REF. 1).
SQ SEQUENCE 2032 AA; 225507 MW; 0256603047FA45EF CRC64;

Query Match 75.0%; Score 6; DB 1; Length 2032;
Best Local Similarity 100.0%; Pred. No. 7 8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRKE 8
Db 874 WKIRKE 879
|||||

RESULT 8
VPU_HV1ZH STANDARD; PRT; 58 AA.
ID VPU_HV1ZH
AC P08806;

DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE VP1 PROTEIN (U ORF PROTEIN) (FRAGMENT).
 GN VP1.
 OS Human immunodeficiency virus type 1 (Zaire H3231 isolate) (HIV-1).
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11692;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89228766; PubMed=2713163;
 RA Srinivasan A., York D., Butler D. Jr., Jannoun-Nasr R., Getchell J.,
 RA McCormick J., Ou C.Y., Myers G., Smith T., Chen E.;
 RT "Molecular characterization of HIV-1 isolated from a serum collected
 RT in 1976: nucleotide sequence comparison to recent isolates and
 RT generation of hybrid HIV";
 RL AIDS Res. Hum. Retroviruses 5:121-129(1989).
 CC -1- FUNCTION: ACTS IN THE DEGRADATION OF CD4 IN THE ENDOPLASMIC
 CC RETICULUM AND IN THE ENHANCEMENT OF VIRION RELEASE FROM THE PLASMA
 CC MEMBRANE OF INFECTED CELLS.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M15896; AAB53947.1; -;
 DR HSSP; P19554; 1VPU.
 DR HIV; M15896; VPU52321.
 DR InterPro: IPR002094; Vpu.
 DR Pfam; PF00558; Vpu; 1.
 KW Transmembrane; AIDS.
 FT NON_TER 1 1
 FT 1
 SQ SEQUENCE 58 AA; 6789 MW; 40EE98A77BB3DACE CRC64;
 Query Match 62.5%; Score 5; DB 1; Length 58;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 KIRKE 8
 DB 9 KIRKE 13
 RESULT 9
 Y414_ARCFU
 ID Y414_ARCFU STANDARD; PRT; 144 AA.
 AC O29833;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN AF0414.
 GN AF0414.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.A., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,

RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing Archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001076; AAB90831.1; -;
 DR TIGR; AF0414; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 144 AA; 16262 MW; 834C76CF066EE59 CRC64;
 Query Match 62.5%; Score 5; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 KIRKE 8
 DB 62 KIRKE 66
 RESULT 10
 YJ218W_YEAST
 ID YJ218W_YEAST STANDARD; PRT; 196 AA.
 AC P40892;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PUTATIVE ACETYLTRANSFERASE IN HXT11-HXT8 INTERGENIC REGION
 DE (EC 2.3.1.-).
 GN YJ218W OR J0224 OR HRA196.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95242842; PubMed=7725802;
 RA Vandenbol M., Durand P., Bolle P.-A., Dion C., Portetelle D.,
 RA Hilger F.;
 RT "Sequence analysis of a 40.2 kb DNA fragment located near the left
 RT telomere of yeast chromosome X.";
 RL Yeast 10:1657-1662(1994).
 CC -1- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NOCL FAMILY OF
 CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
 CC STRONGEST SIMILARITY TO E. COLI LACA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z34098; CAA83992.1; -;
 DR EMBL; Z49493; CAA89515.1; -;
 DR PIR; S45154; S45154.
 DR SGD; S0003754; YJL218W.
 DR InterPro: IPR001451; Hexapep.transf.
 DR Pfam; PF00132; hexapep; 3.
 DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; 1.

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR ENBL: U67470; AAB98113.1; -
DR TIGR; MJ0132; -
DR InterPro; IPR003356; N6_DNA_Mtase.
DR Pfam; PF02384; N6_Mtase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 220 AA; 25766 MW; 710DDAEAC7A47954 CRC64;

Query Match 62.5%; Score 5; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
Db 66 KIRKE 70

RESULT 13
PYRE_COLGR STANDARD; PRT; 233 AA.
AC P35788;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OROTATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.10) (OMPPPASE) (OPRT).
GN PYR1.
OS Colletotrichum graminicola (Anthracnose fungus) (Glomerella
OS graminicola).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
OC Glomerella.
OX NCBI_TaxID=31870;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93052810; PubMed-1435732;
RA Rasmussen J.B., Panaccione D.G., Fang G.C., Hanau R.M.;
RT "The PYR1 gene of the plant pathogenic fungus Colletotrichum
RT graminicola: selection by intraspecific complementation and sequence
RT analysis.";
RL Mol. Gen. Genet. 235:74-80(1992).
CC -|- CATALYTIC ACTIVITY: OROTIDINE-5'-PHOSPHATE + PYROPHOSPHATE =
CC OROTATE + 5-PHOSPHO-ALPHA-D-RIBOSE 1-DIPHOSPHATE.
CC -|- PATHWAY: FIFTH STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -|- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
CC PHOSPHORIBOSYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR ENBL: S47907; AAB24061.1; -
DR PIR; S30118; S30118.
DR HSSP; P08870; ISTO.
DR InterPro; IPR000836; Pribo syltran.
DR Pfam; PF00156; Pribo syltran; 1.
DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
KW Pyrimidine biosynthesis; Transferase; Glycosyltransferase.
FT ACT_SITE 113 113 BY SIMILARITY.
SQ SEQUENCE 233 AA; 25231 MW; 1E6CDB822CA29664 CRC64;

Query Match 62.5%; Score 5; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
Db 150 KIRKE 154

RESULT 14
PYRE_METAN STANDARD; PRT; 234 AA.
AC O42767;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OROTATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.10) (OMPPPASE) (OPRT).
GN URA5.
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae;
OC Metarhizium.
OX NCBI_TaxID=5530;
RN [1]
RP SEQUENCE FROM N.A.
RA Hwang C.-W., Lee D.-K., Kang S.-C.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: OROTIDINE-5'-PHOSPHATE + PYROPHOSPHATE =
CC OROTATE + 5-PHOSPHO-ALPHA-D-RIBOSE 1-DIPHOSPHATE.
CC -|- PATHWAY: FIFTH STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -|- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
CC PHOSPHORIBOSYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR ENBL: AF038545; AAC02431.1; -
DR HSSP; P00495; IORO.
DR InterPro; IPR000836; Pribo syltran.
DR InterPro; IPR002375; Pur_Pyr_pr_transf.
DR Pfam; PF00156; Pribo syltran; 1.
DR PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
KW Transferase; Glycosyltransferase.
FT ACT_SITE 114 114 BY SIMILARITY.
SQ SEQUENCE 234 AA; 25162 MW; 007971EAADD0B253 CRC64;

Query Match 62.5%; Score 5; DB 1; Length 234;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
Db 151 KIRKE 155

RESULT 15
YC34_METJA STANDARD; PRT; 241 AA.
AC Q58631;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1234.
GN MJ1234.

OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.

```

RC SPRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RA MEDLINE=96337999; PubMed=6689087;
RA Sulton G.G., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii."
RT Science 273:1058-1073(1996).
RL -1- SIMILARITY: TO M.JANNASCHII MJ0375.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: U67564; AAB99238.1; -
CC TIGR: MJ1234; -
CC InterPro: IPR002743; DUF57.
CC Pfam: PF01881; DUF57. 1.
CC ProDom: PD008324; DUF57. 1.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 241 AA; 28840 MW; 23AD8F19448CBFC6 CRC64;

Query Match 62.5%; Score 5; DB 1; Length 241;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
Db 209 KIRKE 213
|||||

```

Search completed: January 31, 2002, 13:39:13
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:33 ; Search time 130.99 Seconds
(without alignments)
8.933 Million cell updates/sec

Title: US-08-957-709-69
Perfect score: 8
Sequence: 1 PDWKIRKE 8

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0 473505

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	75.0	418	10 Q9FG12	Q9fg12 arabidopsis
2	6	75.0	2030	13 Q9DDV7	Q9ddv7 xenopus lae
3	6	75.0	2065	13 Q9PT63	Q9pt63 xenopus lae
4	6	75.0	2066	13 Q9DDV8	Q9ddv8 xenopus lae
5	5	62.5	46	6 Q28900	Q28900 bos taurus
6	5	62.5	77	11 Q63078	Q63078 rattus norv
7	5	62.5	83	2 Q9KJ09	Q9kj09 listeria in
8	5	62.5	87	12 Q70888	Q70888 human immun
9	5	62.5	134	10 Q9LSN9	Q9lsn9 arabidopsis
10	5	62.5	137	12 Q69966	Q69966 human immun
11	5	62.5	146	12 Q85278	Q85278 fowlpox vir
12	5	62.5	149	1 Q57923	Q57923 pyrococcus
13	5	62.5	150	2 Q99RN6	Q99rn6 staphylococ
14	5	62.5	167	2 Q9PM97	Q9pm97 campylobact
15	5	62.5	182	1 Q9V1W5	Q9v1w5 pyrococcus
16	5	62.5	195	1 Q9V1A1	Q9v1a1 pyrococcus
17	5	62.5	201	2 Q9F590	Q9f590 agrobacteri
18	5	62.5	202	5 Q96038	Q96038 ciona savig
19	5	62.5	208	4 Q9G2Y3	Q9gz3 homo sapien

20	5	62.5	208	4 Q9BQF7	Q9bqf7 homo sapien
21	5	62.5	210	4 Q9BT06	Q9bt06 homo sapien
22	5	62.5	213	4 Q76080	Q76080 homo sapien
23	5	62.5	213	11 O88878	O88878 mus musculu
24	5	62.5	218	2 Q99RJ6	Q99rj6 staphylococ
25	5	62.5	219	1 O28785	O28785 archaeoglob
26	5	62.5	222	4 Q9NZ51	Q9nzs1 homo sapien
27	5	62.5	222	4 Q9Y3E7	Q9y3e7 homo sapien
28	5	62.5	222	6 Q9GM32	Q9gm32 macaca fasc
29	5	62.5	223	5 Q9VN02	Q9vn02 drosophila
30	5	62.5	223	11 Q9ER79	Q9er79 mus musculu
31	5	62.5	223	11 Q9DCH6	Q9dch6 mus musculu
32	5	62.5	224	11 Q9D7A5	Q9d7a5 mus musculu
33	5	62.5	224	11 Q9D222	Q9d222 mus musculu
34	5	62.5	224	11 Q9CQ10	Q9cq10 mus musculu
35	5	62.5	230	2 Q9PHZ5	Q9phz5 campylobact
36	5	62.5	234	12 Q40171	Q40171 human immun
37	5	62.5	234	12 Q40172	Q40172 human immun
38	5	62.5	234	12 Q40173	Q40173 human immun
39	5	62.5	241	2 O51573	O51573 borrelia bu
40	5	62.5	243	5 Q17922	Q17922 caenorhabdi
41	5	62.5	251	3 Q12592	Q12592 candida mal
42	5	62.5	252	2 Q9X5L0	Q9x5l0 agrobacteri
43	5	62.5	256	5 Q9V3G1	Q9v3g1 drosophila
44	5	62.5	257	2 Q66964	Q66964 aquifex aeo
45	5	62.5	269	1 Q29219	Q29219 archaeoglob

ALIGNMENTS

RESULT 1					
Q9FG12	ID	Q9FG12	PRELIMINARY;	PRT;	418 AA.
AC	Q9FG12				
DT	01-MAR-2001	(Tremblrel. 16, Created)			
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)			
DT	01-JUN-2001	(Tremblrel. 17, Last annotation update)			
DE	GLYCOLATE OXIDASE SUBUNIT D-LIKE, D-LACTATE DEHYDROGENASE-LIKE.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=COLUMBIA;				
RA	Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,				
RA	Tabata S.;				
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. XI.";				
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AP002543; BAB11407.1;				
DR	InterPro; IPR001575; OX1G_FAD_bind.				
DR	Pfam; PF01565; FAD_binding_4; 1.				
SQ	SEQUENCE 418 AA; 45807 MW; D8C84FC0C1E60FA2 CRC64;				

Query Match	75.0%;	Score 6;	DB 10;	Length 418;
Best Local Similarity	100.0%;	Pred. No. 19;		
Matches 6;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	3 WKIRKE 8			
Db	272 WKIRKE 277			
RESULT 2				
Q9DDV7	ID	Q9DDV7	PRELIMINARY;	PRT; 2030 AA.
AC	Q9DDV7;			
DT	01-MAR-2001	(Tremblrel. 16, Created)		
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)		
DT	01-JUN-2001	(Tremblrel. 17, Last annotation update)		

DE MICROTUBULE ASSOCIATED PROTEIN XMAP215 ISOFORM 2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20545155; PubMed=11093249;
RA Becker B.E., Gard D.L.;
RT "Multiple isoforms of the high molecular weight microtubule associated
protein XMAP215 are expressed during development in Xenopus.";
RL Cell Motil. Cytoskeleton 47:282-295(2000).
DR EMBL; AF274064; AAG34915.1; -;
DR InterPro; IPR000357; HEAT_REPEAT.
DR PROSITE; PS00077; HEAT_REPEAT.
DR VARIANT 821 821 T -> S.
FT VARIANT 1103 1105 PPP -> APT.
FT SEQUENCE 2030 AA; 225007 MW; 6AA2B0AA98955BD6 CRC64;
SQ

Query Match 75.08; Score 6; DB 13; Length 2030;
Best Local Similarity 100.0%; Pred. No. 65; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 WKIRKE 8
Db 871 WKIRKE 876

RESULT 3
ID Q9PT63 PRELIMINARY; PRT; 2065 AA.
AC Q9PT63;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MICROTUBULE ASSOCIATED PROTEIN 215 KDA (XMAP215).
GN XMAP215.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20089046; PubMed=10620801;
RA Tournibize R., Popov A., Kinoshita K., Ashford A.J., Rybina S.,
RA Pozniakovskiy A., Mayer T.U., Walczak C.E., Karsenti E., Hyman A.A.;
RT "Control of microtubule dynamics by the antagonistic activities of
XMAP215 and XKCM1 in xenopus egg extracts.";
RL Nat. Cell Biol. 2:13-19(2000).
DR EMBL; AJ251130; CAB61894.1; -;
DR InterPro; IPR000357; HEAT_REPEAT.
DR PROSITE; PS00077; HEAT_REPEAT.
SQ SEQUENCE 2065 AA; 228388 MW; 4C879F481E86D5CB CRC64;

Query Match 75.08; Score 6; DB 13; Length 2065;
Best Local Similarity 100.0%; Pred. No. 66; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 WKIRKE 8
Db 870 WKIRKE 875

RESULT 4
ID Q9DDV8 PRELIMINARY; PRT; 2066 AA.
AC Q9DDV8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MICROTUBULE ASSOCIATED PROTEIN XMAP215 ISOFORM M.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20545155; PubMed=11093249;
RA Becker B.E., Gard D.L.;
RT "Multiple isoforms of the high molecular weight microtubule associated
protein XMAP215 are expressed during development in Xenopus.";
RL Cell Motil. Cytoskeleton 47:282-295(2000).
DR EMBL; AF274063; AAG34914.1; -;
DR InterPro; IPR000357; HEAT_REPEAT.
DR PROSITE; PS00077; HEAT_REPEAT.
DR VARIANT 821 821 T -> S.
FT VARIANT 1103 1105 PPP -> APT.
FT SEQUENCE 2066 AA; 228516 MW; 4F5CA6E4EA27345F CRC64;
SQ

Query Match 75.08; Score 6; DB 13; Length 2066;
Best Local Similarity 100.0%; Pred. No. 66; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 WKIRKE 8
Db 871 WKIRKE 876

RESULT 5
ID Q28900 PRELIMINARY; PRT; 46 AA.
AC Q28900;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE UBIQUITIN HOMOLOG (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95279954; PubMed=7713329;
RA Shima D.T., Saunders K.B., Gougos A., D'Amore P.A.;
RT "Alterations in gene expression associated with changes in the state
of endothelial differentiation.";
RL Differentiation 58:217-226(1995).
DR EMBL; S77733; AAB34029.1; -;
DR InterPro; IPR00058; Znf_AN1.
DR Pfam; PF01428; znf-AN1.1.
DR SMART; SM00154; Znf_AN1.1.
FT NON_TER
FT SEQUENCE 46 AA; 5309 MW; FE274FB6CD907A0D CRC64;

Query Match 62.58; Score 5; DB 6; Length 46;
Best Local Similarity 100.0%; Pred. No. 45; Indels 0; Gaps 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
Db 30 KIRKE 34

RESULT 6
ID Q63078 PRELIMINARY; PRT; 77 AA.
AC Q63078;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CSA COMPLEMENT COMPONENT PROTEIN (FRAGMENT).
 GN CSA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LEWIS; TISSUE=LIVER;
 RX MEDLINE=9723624; PubMed=9116048;
 RA Rothermel E., Rolf O., Goetze O., Zwierner J.;
 RT "Nucleotide and corrected amino acid sequence of the functional
 RT recombinant rat anaphylatoxin CSA."
 RL Biochim. Biophys. Acta 1351:9-12(1997).
 DR EMBL; X91892; CAA62994.1; .
 DR HSP; P01031; 1KJS
 DR InterPro; IPR000020; Anaphylatoxin.
 DR InterPro; IPR001840; Anaphylatoxn.
 DR Pfam; PF01821; ANATO; 1.
 DR PRINTS; PRO0004; ANAPHYLATOXN.
 DR SMART; SM00104; ANATO; 1.
 DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
 DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
 FT NON_TER 1 1
 FT NON_TER 77 77
 FT NON_TER 77 77
 SQ SEQUENCE 77 AA; 8981 MW; 14141F41CC38BD28 CRC64;

Query Match 62.5%; Score 5; DB 11; Length 77;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
 Db 63 KIRKE 67

RESULT 7
 Q9KJU9 PRELIMINARY; PRT; 83 AA.
 AC Q9KJU9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE RHO FACTOR (FRAGMENT).
 GN RHO.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=F8596;
 RA Lan Z., Kathariou S.;
 RT "Region homologous to teichoic acid glycosylation gene gtCA in
 RT Listeria innocua strain F8596."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF160251; AAF80387.1; .
 FT NON_TER 1 1
 FT NON_TER 83 AA; 10155 MW; 818FF940EA62F823 CRC64;

Query Match 62.5%; Score 5; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 WKIRK 7
 Db 40 WKIRK 44

RESULT 8
 Q70888 PRELIMINARY; PRT; 87 AA.
 AC Q70888;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ENVELOPE GLYCOPROTEIN V1V2 REGION (FRAGMENT).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=003;
 RA Cornelissen M., Goudsmit J.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=003;
 RA McEvilly M.M.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U13479; AAA73671.1; .
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR KW Envelope protein.
 FT NON_TER 1 1
 FT NON_TER 87 87
 FT NON_TER 87 87
 SQ SEQUENCE 87 AA; 9786 MW; EAIAACC763C225A8 CRC64;

Query Match 62.5%; Score 5; DB 12; Length 87;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KIRKE 8
 Db 54 KIRKE 58

RESULT 9
 Q9LSN9 PRELIMINARY; PRT; 134 AA.
 AC Q9LSN9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE SIMILARITY TO PAPAYER NIDUCAULE SELF-INCOMPATIBILITY PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakanura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty Pl and TAC
 RT clones."
 RL DNA Res. 7:131-135(2000).
 DR EMBL; AB026636; BAA94986.1; .
 SQ SEQUENCE 134 AA; 15641 MW; F6E8BCE61B5D588E CRC64;

Query Match 62.5%; Score 5; DB 10; Length 134;
 Best Local Similarity 100.0%; Pred. No. 1e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 WKIRK 7

Db 110 WKIRK 114

RESULT 10

Q69966 Q69966 PRELIMINARY; PRT; 137 AA.

AC Q69966; 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE SH750 (ENV) (FRAGMENT).

GN ENV.

OS Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SH750;

RX MEDLINE=96157217; PubMed=8573385;

RA Becker M.L., De Jager G., Becker W.B.;

RT "Analysis of partial gag and env gene sequences of HIV type 1 strains

from southern Africa.";

RL AIDS Res. Hum. Retroviruses 11:1265-1267(1995).

DR EMBL; U07017; AAB01036.1; -.

DR InterPro; IPR000777; GP120.

DR Pfam; PF00516; GP120; 1.

FT NON_TER 137 137

SQ SEQUENCE 137 AA; 14933 MW; 7B95641B8FCF096D CRC64;

Query Match

Best Local Similarity 62.5%; Score 5; DB 12; Length 137;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8

Db 62 KIRKE 66

RESULT 11

Q85278 Q85278 PRELIMINARY; PRT; 146 AA.

AC Q85278; 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE MAJOR ENVELOPE ANTIGEN (P43K) (FRAGMENT).

OS Fowlpox virus.

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

OX Avipoxvirus.

OX NCBI_TaxID=10261;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91021027; PubMed=2219722;

RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,

Paoletti E.;

RT "The complete DNA sequence of vaccinia virus.";

RL Virology 179:247-266(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93079881; PubMed=1333124;

RA Calvert J.G., Ogawa R., Yanagida N., Nazerian K.;

RT "Identification and functional analysis of the fowlpox virus homolog

of the vaccinia virus p37K major envelope antigen gene.";

RL Virology 191:783-792(1992).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=93139784; PubMed=8380837;

RA Ogawa R., Yanagida N., Nazerian K., Calvert J.G.;

RT "Insertional inactivation of a fowlpox virus homologue of the vaccinia

virus F12L gene inhibits the release of enveloped virions.";

RL J. Gen. Virol. 74:55-64(1993).

DR EMBL; M88587; AAA43818.1; -.

FT NON_TER 146 146

SQ SEQUENCE 146 AA; 17065 MW; 1BF8FE96F824A32C CRC64;

Query Match

Best Local Similarity 62.5%; Score 5; DB 12; Length 146;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8

Db 71 KIRKE 75

RESULT 12

Q57923 Q57923 PRELIMINARY; PRT; 149 AA.

AC Q57923; 01-AUG-1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE 149AA LONG HYPOTHETICAL MOLYBDOPTERINCONVERTING FACTOR, SUBUNIT.

GN PH0184.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI_TaxID=53953;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,

Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,

Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

Masuchi Y., Shizuya H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyper-

thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RL DNA Res. 5:55-76(1998).

DR EMBL; AP000001; BAA29253.1; -.

DR InterPro; IPR003448; MoeA.

DR InterPro; IPR000531; TonB_boxC.

DR Pfam; PF02391; MoeA; 1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.

KW Complete Proteome.

SQ SEQUENCE 149 AA; 17105 MW; BCEAC6734DE299D5 CRC64;

Query Match

Best Local Similarity 62.5%; Score 5; DB 1; Length 149;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8

Db 66 KIRKE 70

RESULT 13

Q99RN6 Q99RN6 PRELIMINARY; PRT; 150 AA.

AC Q99RN6;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN SA2181.

GN SA2181.

OS Staphylococcus aureus subsp. aureus N315.

OC Bacteria; Firmicutes; Bacillus/clostridium group;

OC Bacillus/staphylococcus group; Staphylococcus.

```
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA Matsunaru H., Maruyama A., Murakami Y., Hosoyama A., Mizutani-Ui Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003137; BAB43483.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 150 AA; 17150 MW; 006D965569DFCCDD CRC64;

Query Match 62.5%; Score 5; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
Db 18 KIRKE 22

RESULT 14
Q9PM97 PRELIMINARY; PRT; 167 AA.
AC
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE NADH DEHYDROGENASE I CHAIN B (EC 1.6.5.3).
GN NUOB OR CJ1578.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTC 111168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; ALI39079; CAB73566.1; -.
DR InterPro; IPR002096; Complex1_20kd.
DR Pfam; PF01058; oxidored_g6; 1.
KW Complete proteome.
SQ SEQUENCE 167 AA; 18682 MW; D4E83BDE27881BD CRC64;

Query Match 62.5%; Score 5; DB 2; Length 167;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
Db 151 KIRKE 155

RESULT 15
Q9V1W5 PRELIMINARY; PRT; 192 AA.
ID Q9V1W5
AC Q9V1W5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
```

```
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 21.6 KDA PROTEIN.
GN PAB0211.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248284; CAB49233.1; -.
DR HSSP; P03036; 2CRO.
DR InterPro; IPR000644; CBS.
DR InterPro; IPR001387; HTH_3.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF01381; HTH_3; 1.
DR SMART; SM00116; CBS; 2.
DR SMART; SM00530; HTH_XRE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 192 AA; 21645 MW; E0E11E4F88F7A7AA CRC64;
```

```
Query Match 62.5%; Score 5; DB 1; Length 192;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KIRKE 8
Db 15 KIRKE 19

Search completed: January 31, 2002, 13:37:35
Job time: 160 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:04 ; Search time 140.03 Seconds
(without alignments)
82.521 Million cell updates/sec

Title: US-08-957-709-71

Perfect score: 156

Sequence: 1 MLLPDWKIRKEILIEPFSEE.....PYRNGYOGSTRLAFSKRKL 156

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	156	100.0	156	19	AAW72847 Polymerase enhanci
2	14	9.0	15	19	AAW72845 Polymerase enhanci
3	11	7.1	14	19	AAW72849 Pyrococcus furios
4	8	5.1	8	19	AAW72846 Polymerase enhanci
5	7	4.5	15	19	AAW72870 Polymerase enhanci
6	7	4.5	153	21	AA1981410 Bordetella pertuss
7	7	4.5	162	17	AA196207 Attachment-invasio
8	7	4.5	164	21	AA196155 Arabidopsis thalia
9	7	4.5	164	22	AA196292 S. epidermidis ope
10	7	4.5	171	21	AA196154 Arabidopsis thalia
11	7	4.5	198	21	AA196153 Arabidopsis thalia

12	7	4.5	250	21	AA196156 Virulence gene pro
13	7	4.5	362	17	AA196024 Human Protease and
14	7	4.5	568	17	AA196210 AIL protein-maltos
15	7	4.5	607	22	AA196087 Wheat Nph2-1 prote
16	7	4.5	609	22	AA196088 Wheat Nph2-2 prote
17	7	4.5	621	21	AA196271 Maize NPRI transla
18	7	4.5	621	21	AA196272 Maize NPRI transla
19	7	4.5	782	22	AA196302 Human protein sequ
20	7	4.5	2482	16	AA196286 Human mitosin. Ho
21	7	4.5	2482	19	AA196396 Human mitosin amin
22	7	4.5	3248	17	AA1969795 Kinetochore protei
23	6	3.8	8	20	AA196967 Non-crosslinked pr
24	6	3.8	9	11	AA1965947 Somatic immunoglob
25	6	3.8	14	21	AA1966872 T cell antigen rec
26	6	3.8	17	22	AA1968013 Smad interacting p
27	6	3.8	19	22	AA1963385 Trichoderma harzia
28	6	3.8	22	21	AA1964453 C-terminal fragm
29	6	3.8	24	22	AA1968022 Smad interacting p
30	6	3.8	36	22	AA19621642 Peptide #8076 enco
31	6	3.8	36	22	AA19637947 Peptide #11984 enc
32	6	3.8	47	21	AA19617889 Arabidopsis thalia
33	6	3.8	48	21	AA19627581 Human secreted pro
34	6	3.8	52	20	AA1960464 Human normal bladd
35	6	3.8	55	20	AA1960052 Peptide sequence d
36	6	3.8	61	21	AA1963285 Arabidopsis thalia
37	6	3.8	75	21	AA19608097 Arabidopsis thalia
38	6	3.8	76	21	AA19600402 Human secreted pro
39	6	3.8	77	21	AA19634739 Arabidopsis thalia
40	6	3.8	78	21	AA19611055 Arabidopsis thalia
41	6	3.8	81	20	AA19637542 Amino acid sequenc
42	6	3.8	86	20	AA19636015 Extended human sec
43	6	3.8	86	20	AA19635912 Extended human sec
44	6	3.8	90	22	AA19674034 Human colon cancer
45	6	3.8	93	22	AA19679289 Corynebacterium gl

ALIGNMENTS

RESULT 1
AAW72847
ID AAW72847 standard; Protein; 156 AA.
XX
AC AAW72847;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor p45 (dUTPase) component.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
Key Location/Qualifiers
FH Peptide 2..15
FT /note= "N-terminal peptide used to generate
FT primers"
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX
PA (STRA-) STRATAGENE.
XX
PI Hansen CJ, Hogrefe H;
XX
DR WPI; 1998-542284/46.

XX Hansen CJ, Hogrefe H;
 XX WPI; 1998-542284/46.
 XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 PT use in amplification, sequencing and replication
 XX Claim 71; Page 47; 161pp; English.
 XX This is the uridine-binding motif of the P45 component (see
 CC AAW72847) of the polymerase enhancing factor (PEF) of *Pyrococcus*
 CC furiosus DSM 3638. P45 has been identified as a dUTPase that
 CC has polymerase enhancing activity. Sequences are provided (see
 CC AAW72849-57) of the uridine-binding motifs of dUTPases and dCTP
 CC deaminases of *P. furiosus*, *Methanococcus jannaschii*, *Desulfurolobus*
 CC ambivalens, *Escherichia coli*, yeast, human and herpesvirus; a
 CC consensus (see AAW72848) is also provided. A claimed method of
 CC enhancing a nucleic acid polymerase reaction comprises performing
 CC the reaction in the presence of one or more of the following: a
 CC PEF, a dUTPase, a protein that turns over dUTP and a protein
 CC having one or more of the sequences provided in AAW72848-57. A
 CC claimed protein having PEF activity comprises one or more of
 CC sequences given in AAW72848-57. Kits are provided for replicating
 CC nucleic acids, for site-directed mutagenesis, for nucleic acid
 CC sequencing or for amplification (preferably PCR or RT-PCR).
 XX Sequence 14 AA;
 SQ

Query Match 7.1%; Score 11; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.00036;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 91 FAWVDPGWDGN 101
 Db 1 fawvdpwdgn 11
 |||||

RESULT 4
 AAW72846
 ID AAW72846 standard; Peptide; 8 AA.
 AC AAW72846;
 XX 01-MAR-1999 (first entry)
 XX Polymerase enhancing factor P45 component N-terminal peptide.
 XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 KW amplification; sequencing; replication.
 XX *Pyrococcus furiosus* strain DSM 3638.
 OS WO9842860-A1.
 XX 01-OCT-1998.
 XX 20-MAR-1998; 98WO-US05497.
 XX 24-OCT-1997; 97US-0957709.
 PR 21-MAR-1997; 97US-0822774.
 XX (STRA-) STRATAGENE.
 XX Hansen CJ, Hogrefe H;
 XX WPI; 1998-542284/46.
 XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 PT use in amplification, sequencing and replication

XX Claim 17; Page 33; 161pp; English.
 XX This peptide comprises an N-terminal peptide of the P45 component
 CC of the polymerase enhancing factor (PEF) of *Pyrococcus furiosus*
 CC DSM 3638. It has been obtained from isolated P45. PCR primers
 CC (see AAV63861-64) based on the peptide were used to amplify P45 DNA
 CC (see AAV63860). A full-length P45 sequence is provided in AAW72847.
 CC PEF, a predominant component of which is P45 protein, enhances the
 CC activity of *P. furiosus* DNA polymerase, thereby providing
 CC replication products of greater length and purity. P45 is a
 CC dUTPase and possesses polymerase enhancing activity. The invention
 CC provides novel extracts, proteins and complexes that improve the
 CC polymerisation activity of nucleic acid polymerases. These
 CC complexes may include proteins including the P45 N-terminal
 CC peptide. Also included are methods for identifying compositions
 CC with polymerase enhancing activity, for purifying and using these
 CC compositions, and specific extracts, proteins and complexes that
 CC function to enhance polymerase activity. Nucleic acid polymerase
 CC reactions can be enhanced (claimed) by mixing a nucleic acid
 CC template, at least 1 polymerase and a composition having polymerase
 CC enhancing activity. Kits are provided for replicating nucleic
 CC acids. The kits can be used in site-directed mutagenesis, nucleic
 CC acid sequencing or amplification (preferably PCR or RT-PCR).
 XX Sequence 8 AA;
 SQ

Query Match 5.1%; Score 8; DB 19; Length 8;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 PDWKIRKE 11
 Db 1 pdwkirke 8
 |||||

RESULT 5
 AAW72870
 ID AAW72870 standard; Peptide; 15 AA.
 AC AAW72870;
 XX 01-MAR-1999 (first entry)
 XX Polymerase enhancing factor P45 protein N-terminal peptide.
 XX Polymerase enhancing factor; PEF; DNA polymerase; dUTPase; PCR;
 KW amplification; sequencing; replication.
 XX *Pyrococcus furiosus* strain DSM 3638.
 OS
 XX Key Location/Qualifiers
 FT Misc-difference 2 /label= Leu, Tyr
 FT Misc-difference 3 /label= Leu, Val
 FT Misc-difference 4 /label= Arg, Pro
 XX WO9842860-A1.
 XX 01-OCT-1998.
 XX 20-MAR-1998; 98WO-US05497.
 XX 24-OCT-1997; 97US-0957709.
 PR 21-MAR-1997; 97US-0822774.
 XX (STRA-) STRATAGENE.
 XX Hansen CJ, Hogrefe H;
 XX PI
 XX

DR WPI; 1998-542284/46.
 XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 PT use in amplification, sequencing and replication
 XX
 XX Example 5; Page 33; 161pp; English.
 XX This is an N-terminal peptide of the P45 component of the
 CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.
 CC PEF, the predominant components of which are P50 (see AAW72844) and
 CC P45 (see AAW72847), enhances the activity of *P. furiosus* DNA
 CC polymerase. P45 functions as a dUTPase. The invention provides
 CC novel extracts, proteins and complexes that improve the
 CC polymerisation activity of nucleic acid polymerases. These can be
 CC used to improve nucleic acid replication, polymerisation and
 CC amplification (especially in PCR or RT-PCR).
 XX Sequence 15 AA;
 SQ

Query Match 4.5%; Score 7; DB 19; Length 15;
 Best Local Similarity 100.0%; Pred. No. 4.2; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;
 QY 9 RKEILIE 15
 Db | | | | |
 9 rkeillie 15

RESULT 6
 AAB14140
 ID AAB14140 standard; Protein; 153 AA.
 AC AAB14140;
 XX
 XX 02-FEB-2001 (first entry)
 DT
 DE Bordetella pertussis class II gene protein Orf9.
 XX Orf9; bacterial infection; anti-bacterial; vaccine; whooping cough;
 KW type III secretion system; virulence factor; pathogenicity island.
 KW Bordetella pertussis.
 OS
 XX WO200037493-A2.
 PN
 XX 29-JUN-2000.
 PD
 XX 21-DEC-1999; 99WO-EPI0297.
 PF
 XX 21-DEC-1998; 98GB-0028217.
 PR
 XX (ULBR) UNIV LIBRE BRUXELLES.
 PA
 XX Bollen A, Fauconnier A, Godfroid E;
 PI
 XX WPI; 2000-452178/39.
 XX
 XX N-PSDB: AAA64878, AAA64890.
 DR
 XX Novel polypeptides derived from Bordetella pertussis, useful for
 PT treating and diagnosing Bordetella infection -
 PT
 XX Claim 1; Pages 138-139; 165pp; English.
 PS
 XX Bordetella pertussis possesses a type III secretion system. Type III
 CC secretion systems allow bacteria to target virulence factors directly at
 CC host cells. The present sequence is the Orf9 protein of *B. pertussis*.
 CC The present protein is encoded by a Class II type gene and is an
 CC effector protein involved in the type III secretion system of *B.*
 CC pertussis i.e. a Bordetella pathogenicity protein. The gene of the
 CC present protein is located within a pathogenicity island (see AAA64890).
 CC A pathogenicity island is a compact, distinct genetic unit carrying

CC virulence genes. The present protein may be used to treat or diagnose *B.*
 CC pertussis infection, e.g. as a vaccine. Whooping cough is a disease
 CC caused by infection by *B. pertussis*.
 XX Sequence 153 AA;
 SQ

Query Match 4.5%; Score 7; DB 21; Length 153;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 57 YALILTL 63
 Db | | | | |
 86 yaliltl 92

RESULT 7
 AAR96207
 ID AAR96207 standard; Protein; 162 AA.
 AC AAR96207;
 XX
 XX 24-AUG-1996 (first entry)
 DT
 DE Attachment-invasion-locus protein.
 XX
 XX Attachment-invasion-locus protein; drug delivery;
 KW gastrointestinal membrane; transcytosis; bioavailability;
 KW fusion protein; enterocyte; Peyer's patch M-cell.
 XX
 XX *Yersinia enterocolitica* strain 8081c.
 OS
 XX Key Location/Qualifiers
 FH Peptide 19..34
 FT /note= "Peptide from extracellular loop-1"
 FT Peptide 58..76
 FT /note= "Peptide from extracellular loop-2"
 FT Peptide 102..119
 FT /note= "Peptide from extracellular loop-3"
 FT Peptide 145..152
 FT /note= "Peptide from extracellular loop-4"
 XX
 XX WO9613250-A1.
 PN
 XX 09-MAY-1996.
 PD
 XX 20-OCT-1995; 95WO-US13749.
 PF
 XX 27-OCT-1994; 94US-0331393.
 PR
 XX (AMGE-) AMGEN INC.
 PA
 XX Habberfield AD, Jensen-Pippo K;
 PI
 XX WPI; 1996-251447/25.
 DR N-PSDB: AAT27536.
 DR
 XX Therapeutic delivery system utilising bacterial invasion protein - is
 PT not readily degraded in the gut, enhances systemic bio-availability
 PT of therapeutic agents
 PT
 XX Example 1; Fig 2; 110pp; English.
 PS
 XX The sequence represents a bacterial attachment-invasion-locus (AIL)
 CC protein, which may be complexed with a therapeutic agent to
 CC transport the agent across the gastrointestinal membrane barrier by
 CC transcytosis to increase bioavailability 5- to 100-fold. The agent
 CC and bacterial protein may be linked via a degradable peptide
 CC sequence as a fusion protein. The receptor binding region of the
 CC AIL protein involves all or some of the regions from the 4
 CC extracellular loops. This region retains the binding affinity of
 CC the protein, and may be used alone or as part of a fusion protein
 CC for drug delivery. The protein may be fused with maltose binding

CC protein (AAR96208) to form a fusion protein (AAR96209) which may be
CC purified easily by amylose affinity chromatography. The delivery
CC system allows improved transport across enterocytes and Peyer's
CC patch M-cells. The system is not prone to degradation in the gut or
CC early release of biologically active material, and eliminates the
CC need for parenteral administration.
XX
SQ Sequence 162 AA;

Query Match 4.58; Score 7; DB 17; Length 162;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92

|||||||

Db 49 gvigsfa 55

RESULT 8

AAG46155

ID AAG46155 standard; Protein; 164 AA.

XX AC AAG46155;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58037.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

```
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147192.
PR 07-AUG-1999; 99US-0147260.
PR 08-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 10-AUG-1999; 99US-0147493.
PR 11-AUG-1999; 99US-0147935.
PR 12-AUG-1999; 99US-0148171.
PR 13-AUG-1999; 99US-0148319.
PR 14-AUG-1999; 99US-0148341.
PR 15-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 19-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 21-AUG-1999; 99US-0149723.
PR 22-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 24-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 28-AUG-1999; 99US-0151086.
PR 29-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 02-SEP-1999; 99US-0152363.
PR 03-SEP-1999; 99US-0153070.
PR 04-SEP-1999; 99US-0153758.
PR 05-SEP-1999; 99US-0154018.
PR 06-SEP-1999; 99US-0154039.
PR 07-SEP-1999; 99US-0154779.
PR 08-SEP-1999; 99US-0155119.
PR 09-SEP-1999; 99US-0155486.
PR 10-SEP-1999; 99US-0155659.
PR 11-SEP-1999; 99US-0156458.
PR 12-SEP-1999; 99US-0156596.
PR 13-SEP-1999; 99US-0157117.
PR 14-SEP-1999; 99US-0157153.
PR 15-SEP-1999; 99US-0157625.
PR 16-SEP-1999; 99US-0158028.
PR 17-SEP-1999; 99US-0158222.
PR 18-SEP-1999; 99US-0158369.
PR 19-SEP-1999; 99US-0159293.
PR 20-SEP-1999; 99US-0159294.
PR 21-SEP-1999; 99US-0159295.
PR 22-SEP-1999; 99US-0159329.
PR 23-SEP-1999; 99US-0159330.
PR 24-SEP-1999; 99US-0159331.
PR 25-SEP-1999; 99US-0159637.
PR 26-SEP-1999; 99US-0159638.
PR 27-SEP-1999; 99US-0159584.
PR 28-SEP-1999; 99US-0160741.
PR 29-SEP-1999; 99US-0160767.
PR 30-SEP-1999; 99US-0160768.
PR 01-OCT-1999; 99US-0160770.
PR 02-OCT-1999; 99US-0160814.
PR 03-OCT-1999; 99US-0160815.
PR 04-OCT-1999; 99US-0160980.
PR 05-OCT-1999; 99US-0160981.
PR 06-OCT-1999; 99US-0160989.
PR 07-OCT-1999; 99US-0161404.
PR 08-OCT-1999; 99US-0161405.
PR 09-OCT-1999; 99US-0161406.
PR 10-OCT-1999; 99US-0161359.
PR 11-OCT-1999; 99US-0161360.
PR 12-OCT-1999; 99US-0161361.
PR 13-OCT-1999; 99US-0161920.
PR 14-OCT-1999; 99US-0161922.

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.5%; Score 7; DB 21; Length 164;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 REAFVKG 39
Db 113 reafvkg 119
|||||

RESULT 9
AAG82692
ID AAG82692 standard; Protein; 164 AA.
XX
AC AAG82692;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2478.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI: 2001-316495/33.
XX
DR N-PSDB; AAH53542.
XX
PT Nucleic acids encoding polypeptides from 'staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis' -
XX .
XX Claim 18; Page 653; 2188pp; English.
XX
CC AAH53304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO.4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 164 AA;

Query Match 4.5%; Score 7; DB 22; Length 164;
Best Local Similarity 100.0%; Pred. No. 36;
```



```

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151430.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160845.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.58; Score 7; DB 21; Length 171;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 REAFVKG 39
Db 120 reafvkg 126

```

```

RESULT 11
AAG46153
ID AAG46153 standard; Protein; 198 AA.
XX AC AAG46153;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58035.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW Hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132853.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

```



```

XX DE Virulence gene protein #41.
XX KW Virulence gene; antibacterial; vaccine; bacterial infection;
XX KW septicemia; bronchopneumonia; rhinitis; wound infection.
XX OS Pasteurella multocida.
XX PN WO200061724-A2.
XX PD 19-OCT-2000.
XX XX
XX PF 06-APR-2000; 2000WO-US09218.
XX XX
XX PR 09-APR-1999; 99US-0128689.
XX PR 10-SEP-1999; 99US-0153453.
XX XX
XX PA (PHAA ) PHARMACIA & UPJOHN INC.
XX XX
XX PI Lowery DE, Fuller TE, Kennedy MJ;
XX XX
XX DR WPI: 2000-647422/62.
XX DR N-PSDB; AAC79622.
XX XX
XX PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence
XX PT genes, useful as a live attenuated vaccine against bacterial infections
XX PT
XX PS Claim 39; Pages 232-233; 322pp; English.
XX CC
XX CC The family Pasteurellaceae encompasses several pathogens that infect a
XX CC wide variety of animals. The present invention relates to virulence genes
XX CC from Pasteurellaceae. The present invention is a protein encoded by one
XX CC such virulence gene. The virulence genes of the present invention may be
XX CC mutated in order to produce an inactive gene. The inactive virulence gene
XX CC may in turn be used to produce a vaccine, which is useful for treating
XX CC bacterial infections such as septicemias, bronchopneumonias, rhinitis and
XX CC wound infections.
XX CC
XX SQ Sequence 250 AA;

Query Match 4.5%; Score 7; DB 21; Length 250;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 SLAREGV 87
DB 38 slaregv 44

RESULT 13
AAY70024
XX AC AAY70024 standard; Protein: 362 AA.
XX XX
XX DT 05-JUN-2000 (first entry)
XX DE Human Protease and associated protein-18 (PPRG-18).
XX KW
XX KW Protease and associated protein-18; PPRG-18; anti-PPRG antibody;
XX KW diagnosis; treatment; cell proliferative disorder; cancer; cirrhosis;
XX KW arteriosclerosis; atherosclerosis; bursitis; hepatitis; immune disorder;
XX KW AIDS; Addison's disease; adult respiratory distress syndrome; allergy;
XX KW ankylosing spondylitis; amyloidosis; cytostatic; antiarteriosclerotic;
XX KW hepatotrophic; antiinflammatory; virucide; antipsoriatic; anti-HIV;
XX KW antiallergic; immunosuppressive; antidiabetic; antianaemic;
XX KW neuroprotective; human.
XX OS Homo sapiens.
XX XX
XX FH Key . Location/Qualifiers

us-08-957-709-71.rag
XX FT Modified-site 39 /note= "Potential Phosphorylation site"
XX FT Modified-site 40 /note= "Potential Phosphorylation site"
XX FT Modified-site 69 /note= "Potential Phosphorylation site"
XX FT Region 71..108 /note= "Signature sequence of ubiquitin carboxyl-terminal
XX FT 129 hydrolase"
XX FT Modified-site 130 /note= "Potential Phosphorylation site"
XX FT Modified-site 130 /note= "Potential Phosphorylation site"
XX FT Modified-site 166 /note= "Potential Phosphorylation site"
XX FT Modified-site 348 /note= "Potential Phosphorylation site"
XX FT /note= "Potential Phosphorylation site"
XX XX WO200009709-A2.
XX PN
XX XX
XX PD 24-FEB-2000.
XX XX
XX DR 06-AUG-1999; 99WO-US17818.
XX XX
XX PR 10-AUG-1998; 98US-0096114.
XX PR 11-FEB-1999; 99US-0119768.
XX XX
XX PA (INCY-) INCYTE PHARM INC.
XX XX
XX PI Bandman O, Hillman JL, Baughn MR, Azimzai Y, Guegler KJ;
XX PI Corley NC, Yue H, Tang YT, Reddy R, Patterson C, Au-Young J;
XX PI Shih LL, Lu DAM;
XX XX
XX DR WPI: 2000-224346/19.
XX DR N-PSDB; AAZ50935.
XX XX
XX PT New human proteases, useful for diagnosis, treatment and prevention of
XX PT cell proliferative disorders such as atherosclerosis -
XX XX
XX PS Claim 1; Page 95-96; 114pp; English.
XX CC
XX CC The present sequence is that of human protease and associated protein-18
XX CC (PPRG-18), which is expressed in haematopoietic, immune, nervous and
XX CC reproductive tissues. It is encoded by cDNA identified in Incyte clone
XX CC 2820384 derived from BRSTN0T14 cDNA library. Anti-PPRG antibodies can
XX CC be used as therapeutic antagonists. Reagents for diagnosis and monitoring
XX CC diseases and for isolating PPRG. PPRG nucleotide sequence can be used
XX CC as probe or primer for diagnosis and monitoring of PPRG-related
XX CC diseases and gene mapping. PPRG can be used in the treatment of cell
XX CC proliferative disorders like cancer, arteriosclerosis, atherosclerosis,
XX CC bursitis, cirrhosis and hepatitis, and immune disorders like AIDS,
XX CC Addison's disease, adult respiratory distress syndrome, allergies,
XX CC ankylosing spondylitis and amyloidosis.
XX XX
XX SQ Sequence 362 AA;

Query Match 4.5%; Score 7; DB 21; Length 362;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 EAFVKGK 40
DB 273 eafvkgk 279
|||||
|||||

RESULT 14
AAR96210
XX ID AAR96210 standard; Protein: 568 AA.
XX XX
XX AC AAR96210;
XX XX
XX DT 24-AUG-1996 (first entry)

```

XX AIL protein-maltose binding protein fusion protein.
 DE
 XX
 KW Attachment-invasion-locus protein; maltose binding protein;
 KW fusion protein; affinity tail; amylose; affinity chromatography;
 KW purification; drug delivery; gastrointestinal membrane;
 KW transcytosis; bioavailability; enterocyte; Peyers patch M-cell.
 XX
 OS Chimeric: Synthetic;
 OS Chimeric: Versinia enterocolitica.
 XX
 XX
 FH Key Location/Qualifiers
 FT Region 1..392
 FT /note= "Maltose binding protein"
 FT Region 393..568
 FT /note= "Attachment-invasion-locus protein"
 FT Peptide 425..440
 FT /note= "Peptide from extracellular loop-1"
 FT Peptide 464..482
 FT /note= "Peptide from extracellular loop-2"
 FT Peptide 508..525
 FT /note= "Peptide from extracellular loop-3"
 FT Peptide 551..558
 FT /note= "Peptide from extracellular loop-4"
 XX
 XX WO9613250-A1.
 PN
 XX
 XX 09-MAY-1996.
 PD
 XX
 XX 20-OCT-1995; 95WO-US13749.
 PF
 XX
 XX 27-OCT-1994; 94US-0331393.
 PR
 XX
 XX (AMGE-) AMGEN INC.
 PA
 XX
 XX Habberfield AD, Jensen-Pippen K;
 PI
 XX
 XX WPI; 1996-251447/25.
 DR
 XX
 XX Therapeutic delivery system utilising bacterial invasion protein - is
 PT not readily degraded in the gut, enhances systemic bio-availability
 PT of therapeutic agents
 PT
 XX
 PS Example 3; Fig 15; 110pp; English.
 CC
 CC The sequence represents a fusion protein of a Versinia
 CC enterocolitica attachment-invasion-locus (AIL) protein (AAR96207)
 CC with maltose binding protein (MBP, AAR96208). The MBP N-terminal
 CC portion of the fusion protein acts as an affinity tail, allowing
 CC efficient recombinant protein purification by amylose affinity
 CC chromatography. The receptor binding region of the AIL protein
 CC involves all or some of the regions from the 4 extracellular loops.
 CC This region retains the binding affinity of the protein, and may be
 CC used alone or as part of a fusion protein for drug delivery. The
 CC bacterial AIL protein may be complexed with a therapeutic agent to
 CC transport the agent across the gastrointestinal membrane barrier by
 CC transcytosis to increase bioavailability 5- to 100-fold. The
 CC delivery system allows improved transport across enterocytes and
 CC Peyer's patch M-cells. The system is not prone to degradation in the
 CC gut or early release of biologically active material, and
 CC eliminates the need for parenteral administration.
 XX
 XX Sequence 568 AA;

Query Match 4.5%; Score 7; DB 17; Length 568;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 GVIGSFA 92
 |||||
 DB 455 gvigsfa 461

RESULT 15
 AAB48087
 ID AAB48087 standard; Protein; 607 AA.
 XX
 AC AAB48087;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE Wheat Nph2-1 protein sequence.
 XX
 KW Acquired resistance gene; Nph1; Nph2; rice; Nph2-1; Nph2-2; wheat;
 KW plant pathogen; transgenic; disease resistance.
 XX
 OS Triticum aestivum.
 XX
 PN WO200070069-A1.
 PD 23-NOV-2000.
 XX
 PF 12-MAY-2000; 2000WO-US13307.
 XX
 PR 13-MAY-1999; 99US-0133965.
 XX
 PA (MONS) MONSANTO CO.
 XX
 XX Bougri OV, Rommens CMT, Srivastava N, Swords KM;
 PI
 XX WPI; 2001-016244/02.
 DR N-PSDB; AAC84342, AAC84344.
 XX
 XX New acquired resistance genes Nph1 from rice, Oryza sativa, and Nph2-1
 PT and Nph2-2 from wheat, Triticum aestivum, useful for producing
 PT transgenic plants with increased disease resistance
 XX
 PS Claim 19; Page 84-86; 101pp; English.
 XX
 CC The invention relates to acquired resistance genes Nph1 from rice, and
 CC Nph2-1 and Nph2-2 from wheat. The Nph1 and Nph2 polypeptides can be
 CC expressed by standard recombinant methodology. The Nph1 and Nph2
 CC polynucleotides or polypeptides can be used to enhance acquired
 CC resistance in plants (e.g. wheat or rice) to control plant pathogens e.g.
 CC the genes can be introduced to make transgenic plants with increased
 CC disease resistance. The polynucleotides are also useful to produce probes
 CC and primers useful to detect the polynucleotides (to identify transgenic
 CC plants containing an acquired resistance gene) and to isolate similar
 CC sequences e.g. from other species. The polypeptides can be used to make
 CC antibodies useful to monitor protein production e.g. in transgenic
 CC plants. The present sequence represents the amino acid sequence of the
 CC wheat Nph2-1 protein.
 XX
 SQ Sequence 607 AA;

Query Match 4.5%; Score 7; DB 22; Length 607;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 RYCREAF 36
 |||||
 DB 146 rvgreat 152

Search completed: January 31, 2002, 13:18:05
 Job time: 166 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:08 ; Search time 65.13 Seconds
(without alignments)
53.900 Million cell updates/sec

Title: US-08-957-709-71

Perfect score: 156

Sequence: 1 MLLPDWKIRKEILIEPSEE.....PYRNGYOGSTRIAFSKRKKL 156

Scoring table: OLIGO

Gapex 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	54.5	246	4	US-08-822-774-44
2	20	12.8	42	4	US-08-822-774-41
3	14	9.0	17	4	US-08-822-774-11
4	11	7.1	14	4	US-08-822-774-54
5	7	4.5	18	4	US-08-822-774-13
6	7	4.5	27	4	US-08-822-774-38
7	7	4.5	568	5	PCT-US95-13749-5
8	7	4.5	2482	1	US-08-328-254-6
9	7	4.5	3248	5	PCT-US95-16216-1
10	7	4.5	3248	5	PCT-US95-16216-1
11	6	3.8	8	2	US-08-747-137-16
12	6	3.8	253	4	US-09-251-645-3
13	6	3.8	257	1	US-07-781-355-2
14	6	3.8	261	3	US-09-058-489-11
15	6	3.8	283	2	US-08-852-401-4
16	6	3.8	297	1	US-08-534-910B-6
17	6	3.8	297	1	US-08-534-910B-7
18	6	3.8	297	1	US-08-534-910B-8
19	6	3.8	297	1	US-08-534-910B-9
20	6	3.8	297	1	US-08-534-910B-10
21	6	3.8	297	3	US-08-886-466-2
22	6	3.8	297	4	US-09-475-304-2
23	6	3.8	353	1	US-08-229-418-2
24	6	3.8	353	2	US-08-932-761A-2
25	6	3.8	353	4	US-09-307-912-2
26	6	3.8	353	5	PCT-US95-04464-2
27	6	3.8	367	3	US-09-141-047-10

28	6	3.8	410	1	US-08-190-802A-46	Sequence 46, Appl
29	6	3.8	410	4	US-08-477-346-46	Sequence 46, Appl
30	6	3.8	425	1	US-08-190-802A-58	Sequence 58, Appl
31	6	3.8	425	3	US-08-105-454-2	Sequence 2, Appl
32	6	3.8	425	4	US-08-477-346-58	Sequence 58, Appl
33	6	3.8	457	4	US-08-477-928A-33	Sequence 33, Appl
34	6	3.8	463	4	US-08-477-928A-6	Sequence 6, Appl
35	6	3.8	476	4	US-09-522-800-16	Sequence 16, Appl
36	6	3.8	476	4	US-08-924-183-1	Sequence 1, Appl
37	6	3.8	476	4	US-08-924-183-7	Sequence 7, Appl
38	6	3.8	476	4	US-09-488-364-1	Sequence 1, Appl
39	6	3.8	476	4	US-09-488-364-7	Sequence 7, Appl
40	6	3.8	478	4	US-08-477-928A-32	Sequence 32, Appl
41	6	3.8	478	4	US-08-477-928A-34	Sequence 34, Appl
42	6	3.8	480	1	US-08-476-008-49	Sequence 49, Appl
43	6	3.8	480	1	US-08-306-063-49	Sequence 49, Appl
44	6	3.8	480	1	US-08-833-485-49	Sequence 49, Appl
45	6	3.8	480	4	US-09-137-440-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1

US-08-822-774-44
; Sequence 44, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREPE, Holly

; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lendahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES

US-08-822-774-44

Query Match 54.5%; Score 85; DB 4; Length 246;

Best Local Similarity 100.0%; Pred. No. 1.1e-74;

Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 VMGDMKIRSLAREGVIGSFAWDGNDLTLMLYNASNEPVELRYGERFVQIAFIRLE 131

Db 20 VMGDMKIRSLAREGVIGSFAWDGNDLTLMLYNASNEPVELRYGERFVQIAFIRLE 79

OY 132 GPARNYRGNYQGSTRLAFSRKKL 156
| | | | | | | | | | | | | | | | | | | | |
Db 80 GPARNYRGNYQGSTRLAFSRKKL 104

RESULT 2

US-08-822-774-41
; Sequence 41, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESS: Lenahan, P.L.L.C. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-822-774-41

Query Match 12.8% Score 20; DB 4; Length 42;
Best Local Similarity 100.0%; Pred. No. 1.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLLPDWKIRKEILIEPSEE 20
| | | | | | | | | | | | | | | | | | | | |
Db 11 MLLPDWKIRKEILIEPSEE 30

RESULT 3

US-08-822-774-11
; Sequence 11, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESS: Lenahan, P.L.L.C.

STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-822-774-11

Query Match 9.0% Score 14; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 LLPDWKIRKEILIE 15
| | | | | | | | | | | | | | | | | | | | |
Db 4 LLPDWKIRKEILIE 17

RESULT 4

US-08-822-774-54
; Sequence 54, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESS: Lenahan, P.L.L.C. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800

TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-822-774-54

Query Match 7.1%; Score 11; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 FAWVDPGWDGN 101
|||||
DB 1 FAWVDPGWDGN 11

RESULT 5

US-08-822-774-13
Sequence 13, Application US/08822774
Patent No. 6183997
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
TITLE OF INVENTION: and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
ADDRESSEE: Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REFERENCE/DOCKET NUMBER: 36,576
REGISTRATION NUMBER: 1486/43163

TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 18 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-822-774-13

Query Match 4.5%; Score 7; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RKEILIE 15
|||||

Db 12 RKEILIE 18

RESULT 6

US-08-822-774-38
Sequence 38, Application US/08822774
Patent No. 6183997
GENERAL INFORMATION:
APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
TITLE OF INVENTION: and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
ADDRESSEE: Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REFERENCE/DOCKET NUMBER: 36,576
REGISTRATION NUMBER: 1486/43163

TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-822-774-38

Query Match 4.5%; Score 7; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 VEKEGV 50
|||||
DB 15 VEKEGV 21

RESULT 7

PCT-US95-13749-5
Sequence 5, Application PC/TUS9513749
GENERAL INFORMATION:
APPLICANT: Amgen Inc.
TITLE OF INVENTION: COMPOSITIONS FOR INCREASED
TITLE OF INVENTION: BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:

Query Match 4.5%; Score 7; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RKEILIE 15
|||||

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13749
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 568 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US95-13749-5

Query Match 4.5%; Score 7; DB 5; Length 568;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 GVIGSFA 92

Db 455 GVIGSFA 461

|||||

RESULT 8

US-08-328-254-6

; Sequence 6, Application US/08328254

; Patent No. 5710022

; GENERAL INFORMATION:

; APPLICANT: Zhu, Xuellang

; APPLICANT: Lee, Wen-Hwa

; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/328,254

; FILING DATE: 24-Oct-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/141,239

; FILING DATE: 22-OCT-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-CJ 1191

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2482 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-328-254-6

Query Match 4.5%; Score 7; DB 1; Length 2482;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 VEKEGV 50

Db 2054 VEKEGV 2060

|||||

RESULT 9

US-08-353-700-1

; Sequence 1, Application US/08353700

; Patent No. 5595919

; GENERAL INFORMATION:

; APPLICANT: YEN, TIMOTHY J.

; APPLICANT: RATTNER, JEROME B.

; TITLE OF INVENTION: NUCLEIC ACID ENCODING A

; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,

; TITLE OF INVENTION: AND METHODS OF USE

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN

; STREET: 1601 MARKET STREET, SUITE 720

; CITY: PHILADELPHIA

; STATE: PA

; COUNTRY: USA

; ZIP: 19103-2307

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/353,700

; FILING DATE: 09-DEC-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: REED, JANET E.

; REGISTRATION NUMBER: 36,252

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 563-4100

; TELEFAX: (215) 563-4044

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3248 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: HUMAN

US-08-353-700-1

Query Match 4.5%; Score 7; DB 1; Length 3248;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 VEKEGV 50

Db 2782 VEKEGV 2788

|||||

RESULT 10

PCT-US95-16216-1

; Sequence 1, Application PC/TUS9516216

; GENERAL INFORMATION:

; APPLICANT: Yen, Timothy J.

; APPLICANT: Rattner, Jerome B.

; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently

; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 4.5%; Score 7; DB 5; Length 3248;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 VEKEGV 50
Db 2782 VEKEGV 2788

RESULT 11
US-08-747-137-16
Sequence 16, Application US/08/47137
Patent No. 5945033
GENERAL INFORMATION:
APPLICANT: YEN, Richard C.K.
TITLE OF INVENTION: NON-CROSSLINKED PROTEIN PARTICLES FOR
TREATMENT OF INFECTION: THERAPEUTIC AND DIAGNOSTIC USE
NUMBER OF SEQUENCES: 184
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/747,137
FILING DATE: 12-NOV-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,546
FILING DATE: 14-MAR-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/069,831
FILING DATE: 01-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/959,560
FILING DATE: 13-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/641,720
FILING DATE: 15-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 016197-0008400S
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
US-08-747-137-16

Query Match 3.8%; Score 6; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 ALILTL 63
Db 1 ALILTL 6

RESULT 12
US-09-251-645-3
Sequence 3, Application US/09251645
Patent No. 6281413
GENERAL INFORMATION:
APPLICANT: Kramer, Vance C.
APPLICANT: Morgan, Michael K.
APPLICANT: Anderson, Arne R.
APPLICANT: Hart, Hope
APPLICANT: Warren, Gregory W.
APPLICANT: Dunn, Martha
APPLICANT: Chen, Jeng S.
TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
FILE REFERENCE: CGC1963/A
CURRENT APPLICATION NUMBER: US/09/251,645
CURRENT FILING DATE: 1999-02-17
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 3
LENGTH: 253
TYPE: PRT
ORGANISM: Photorhabdus luminescens
US-09-251-645-3

Query Match 3.8%; Score 6; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KIRKEI 12
Db 192 KIRKEI 197

RESULT 13
US-07-781-355-2
Sequence 2, Application US/07781355
Patent No. 5246844
GENERAL INFORMATION:

APPLICANT: No. 5246844ris, Steven J.
APPLICANT: Barbour, Alan G.
TITLE OF INVENTION: VIRULENCE ASSOCIATED PROTEINS IN
TITLE OF INVENTION: BORRELIA BURGDORFERI (Bb)
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/781,355
FILING DATE: 19911022
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: UTSH:162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512-320-7200
TELEFAX: 512-474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-781-355-2

Query Match 3.8%; Score 6; DB 1; Length 257;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 KLIDVE 45
Db 202 KLIDVE 207

RESULT 14
US-09-058-489-11
Sequence 11, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE OF INVENTION: the Y Chromosome
FILE REFERENCE: WHI97-08pa
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 261
TYPE: PRT
ORGANISM: C. acetobutylicum
US-09-058-489-11

Query Match 3.8%; Score 6; DB 3; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 EKEGKV 50
Db 9 EKEGKV 14

RESULT 15
US-08-852-401-4
Sequence 4, Application US/08852401
Patent No. 5976836
GENERAL INFORMATION:
APPLICANT: Weber, J. Mark
APPLICANT: Hessler, Paul E.
APPLICANT: Larsen, Peter E.
APPLICANT: Luu, B. Minh
TITLE OF INVENTION: Methods and Compositions for Enhancing
TITLE OF INVENTION: Erythromycin Production
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rocky, Milnamow & Katz, Ltd.
STREET: 2 Prudential Plaza, 180 N. Stetson, Suite
STREET: 4700
CITY: Chicago
STATE: Illinois
COUNTRY: U.S.A.
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,401
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mueller, Lisa L.
REGISTRATION NUMBER: 38,978
REFERENCE/DOCKET NUMBER: FER2159POO30US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-616-5400
TELEFAX: 312-616-5460
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-852-401-4

Query Match 3.8%; Score 6; DB 2; Length 283;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 LYNASN 111
Db 206 LYNASN 211

Search completed: January 31, 2002, 13:15:09
Job time: 90 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:07:05 ; Search time 33.61 Seconds
(without alignments)
343.809 Million cell updates/sec

Title: US-08-957-709-71

Perfect score: 806

Sequence: 1 MLLPDWKIRKEILIEPFSEE.....PYRNGYQGSTRLAFSKRKL 156

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*

1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	806	100.0	156	19 AAW72847	Polymerase enhanci
2	242	30.0	173	14 AAF51078	Polypeptide encode
3	218	27.0	189	22 AAG92886	C glutamicum prote
4	196	24.3	160	22 AAB80139	Corynebacterium gl
5	181	22.5	188	20 AAW89818	Protein encoded by
6	181	22.5	190	20 AAW89897	Antigen 3 from clu
7	149.5	18.5	190	20 AAY37187	Protein involved i
8	138	17.1	190	20 AAY34994	Protein involved i
9	134.5	16.7	177	20 AA129169	Amino acid sequenc
10	119.5	14.8	1143	21 AAB35755	EIAV pol gene prod
11	112	13.9	181	20 AAY29165	Amino acid sequenc

12	111	13.8	141	16	AAW70144	Human dUTPase prot
13	111	13.8	164	18	AAW30280	Human dUTPase (nuc
14	111	13.8	252	18	AAW30281	Human dUTPase (mit
15	111	13.8	1105	8	AAW71667	Translation produc
16	98.5	12.2	166	21	AAG30341	Arabidopsis thalia
17	98.5	12.2	198	21	AAG30340	Arabidopsis thalia
18	97	12.0	254	20	AAI29166	Amino acid sequenc
19	92	11.4	1124	13	AAI24237	Pol polyepptide of
20	85.5	10.6	1031	19	AAW79062	FIV-NC501 clone JS
21	82.5	10.2	1124	19	AAW53668	FIV PPR clone 34 O
22	81	10.0	154	22	AAG81199	Mycobacterium tube
23	80	9.9	149	22	AAG91840	C glutamicum prote
24	79	9.8	470	22	AAG82360	S. epidermidis ope
25	78.5	9.7	142	18	AAW26421	Swinepox virus Hin
26	78.5	9.7	142	22	AAW68242	Protein encoded by
27	78	9.7	201	19	AAW71498	Helicobacter poly
28	76.5	9.5	14	19	AAW72849	Pyrococcus furios
29	76.5	9.5	154	18	AAW20848	H. pylori cytoplas
30	76.5	9.5	659	17	AAW89265	Heparinase-III. F
31	76	9.4	995	15	AAW60812	Agarase 0107 from
32	75	9.3	206	18	AAW55475	H. pylori ORF 07ap
33	75	9.3	209	18	AAW55266	H. pylori ORF 07ap
34	74	9.2	594	20	AAW17899	Methanococcus jann
35	74	9.2	594	21	AAI52017	M. jannaschii MJ07
36	74	9.2	594	21	AAI51646	M. jannaschii MJ07
37	74	9.2	860	20	AAV41140	Mouse mammary tumo
38	74	9.2	1755	20	AAV41139	Mouse mammary tumo
39	73.5	9.1	169	18	AAW28034	Amino acid sequenc
40	73	9.1	15	19	AAW72845	Polymerase enhanci
41	72	8.9	274	21	AAW28243	Arabidopsis thalia
42	72	8.9	309	21	AAW28242	Arabidopsis thalia
43	72	8.9	315	21	AAW28241	Arabidopsis thalia
44	72	8.9	424	11	AAW07058	Mouse zp3 gene pro
45	71	8.8	372	21	AAW67287	Human zona pelluci

ALIGNMENTS

RESULT 1
AAW72847
ID AAW72847 standard; Protein: 156 AA.
XX
AC AAW72847;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 (dUTPase) component.
XX
KW Polymerase enhancing factor; PEF: dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
FH Key Location/Qualifiers
FT Peptide 2..15
FT /note= "N-terminal peptide used to generate
FT primers"
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
XX
PR 21-MAR-1997; 97US-0822774.
XX
PA (STRA-) STRATAGENE.
XX
PI Hansen CJ, Hogrefe H;
XX
DR WPI; 1998-542284/46.

DR N-PSDB; AAV63860.
 XX Polymerase enhancing factor proteins, extracts and complexes -
 PT improve the polymerisation activity of nucleic acid polymerases, for
 PT use in amplification, sequencing and replication
 XX
 PS Claim 17; Page 43; 161pp; English.
 XX
 CC This is the amino acid sequence of the P45 component of the
 CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.
 CC The sequence is predicted from a DNA sequence (see AAV63860) obtained
 CC from genomic DNA by PCR. P45 and P50 (see AAV72844) are the
 CC predominant components of PEF, which acts to enhance the activity of
 CC *P. furiosus* DNA polymerase. P45 functions as a dUTPase, and can be
 CC used to enhance nucleic acid replication, polymerisation or PCR
 CC reactions. The invention provides novel extracts, proteins and
 CC complexes that improve the polymerisation activity of nucleic acid
 CC polymerases, as well as DNA constructs and antibodies. Also
 CC included are methods for identifying compositions with PEF activity,
 CC for purifying and using these compositions, and specific extracts,
 CC proteins and complexes that function to enhance polymerase activity.
 CC Nucleic acid polymerase reactions can be enhanced (claimed) by
 CC mixing a nucleic acid template, at least 1 polymerase and a
 CC composition having polymerase enhancing activity. Kits are
 CC provided for replicating nucleic acids, for site-directed
 CC mutagenesis, for nucleic acid sequencing or amplification
 CC (preferably PCR or RT-PCR).
 XX Sequence 156 AA;

Query Match 100.0%; Score 806; DB 19; Length 156;
 Best Local Similarity 100.0%; Pred. No. 1.5e-89;
 Matches 156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLPDWKIRKEILLIEPSEESLQAGYDLRVGREAFVKGKLDIVKEGKVIPPVEYALI 60
 Db 1 mllpdwkirkeilliepseeslqagydlrvgreafvkgkldivkegkvippveyali 60
 QY 61 LTLERIKLPDDVMGDMKIRSSLAREGVIGSAFWDPGWDGNLTMLYNASNEPVELRYGE 120
 Db 61 ltleriklpddvmgdmkirsrlaregvigsawdpdgwdgnltmllynasnepvelryge 120
 QY 121 RFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRKL 156
 Db 121 rfvqiafirlegparnpyrgnyqgstrlafskrkl 156

RESULT 2
 AAR51078
 ID AAR51078 standard; Protein; 173 AA.
 XX
 AC AAR51078;
 XX
 DT 27-MAY-1994 (first entry)
 XX
 DE Polypeptide encoded by pDam-L3 plasmid fragment.
 XX
 KW Ligase, thermostable; thermostability; Desulfurolobus ambivalens;
 KW archaeobacteria; detection; ligation; ATP; adenosine triphosphate;
 XX
 PN DE4217134-A.
 XX
 XX 25-NOV-1993.
 XX
 XX 23-MAY-1992; 92DE-4217134.
 XX
 XX 23-MAY-1992; 92DE-4217134.
 XX
 PA (BOEF) BOEHRINGER MANNHEIM GMBH.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Jarsch M, Kaletta C, Kessler C, Kletzin A, Rueger R;

XX WPI; 1993-378402/48.
 DR N-PSDB; AAV62300.
 XX
 PT Thermostable ligase from archaeobacteria - and DNA coding for it,
 PT use in amplification, sequencing and replication
 XX
 PS Example 3; Figure 1; 26pp; German.
 XX
 CC A sequence encoding a thermostable DNA ligase was cloned into the
 CC plasmid pDam-L3. The DNA ligase is ATP dependent and is useful for
 CC detecting nucleic acids by hybridising two oligonucleotides with
 CC adjacent sequences of the target nucleic acid, ligating the two
 CC oligonucleotides (using the ligase) and detecting the ligation
 CC product. Analysis of the fragment revealed three other open reading
 CC frames (ORF's). This is a polypeptide encoded by ORF3
 XX Sequence 173 AA;

Query Match 30.0%; Score 242; DB 14; Length 173;
 Best Local Similarity 35.9%; Pred. No. 3.6e-21;
 Matches 56; Conservative 30; Mismatches 56; Indels 14; Gaps 3;
 QY 12 ILIEPSEESLQAGYDLRVG-----REAFVKGK----LIDVEKEGVIPPVEYA 58
 Db 17 ivisplqtqitirengvdrlrvgeiarfktdelyedgkdrfsfyelekgedfiiypnehv 76
 QY 59 LIITLERIKLPDDVMGDMKIRSSLAREGVIGSAFWDPGWDGNLTMLYNASNEPVELRY 118
 Db 77 llvteeyvklpndvmafnlrsslafarlgflfvpptivdagfegqltiavl-gsafpvkikr 135
 QY 119 GERFVQIAFIRLEGPARNPYRGNYQGSTRLAFSKRK 154
 Db 136 gtrflhlifartitpvenpyhgkyqggvgvclpkfk 171

RESULT 3
 AAG92886
 ID AAG92886 standard; Protein; 189 AA.
 XX
 AC AAG92886;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 6640.
 XX
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EPI108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX WPI: 2001-376931/40.
 DR N-PSDB; AAV68105.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene


```

Query Match      24.3%; Score 196; DB 22; Length 160;
Best Local Similarity 32.9%; Pred. No. 1.2e-15;
Matches 52; Conservative 29; Mismatches 53; Indels 24; Gaps 4;

QY 1 MLLPWKIRK-----EILIEPFSEESLOPAGYDLRVGR--EAFVKGK-----40
DB 1 vllsdrdrksldagldgldpdaelqpsvvdvrmdrfvrnnskythidpknqdel 60
QY 41 --LIDVEKEGVIPPREYALITLERIKLPDDVMDGMKIRSSLAREGVI--GSFAWVDP 96
DB 61 tslvedegefvlpbgfvlstlektlphlagrlegkssigrllthstgafidp 120
QY 97 GWDGNLTMLYNASNEPVELRYGERFVQIATRLREGPA 134
DB 121 gfsyvitlslsnvanpltlwpgmkvgqlafqmsppa 158

RESULT 5
AAW89818
ID AAW89818 standard; Protein: 188 AA.
XX
AC AAW89818;
XX
XX
DT 18-FEB-1999 (first entry)
DE Protein encoded by clone B17 ORF4.
XX
KW Antigen: immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
OS Helicobacter pylori.
XX
PN W09849314-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-US08487.
XX
PR 14-OCT-1997; 97US-0061958.
XX
PR 25-APR-1997; 97US-0045107.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Chow TP, Fry KE, Lim MY, McAtee CP;
XX
DR WPI; 1999-009433/01.
XX
DR N-PSDB; AAV90555.
XX
XX
New Helicobacter pylori antigens and related nucleic acid sequences
- useful in serological diagnosis and protective vaccines, providing
long-lasting immune response
Claim 15; Page 104-105; 402pp; English.
XX
XX
The present sequence represents a Helicobacter pylori antigenic protein
that is characterised by immunoreactivity with H. pylori-positive
antisera. The proteins are highly immunogenic and induce a long-lasting
immune response that persists even after antimicrobial treatment. In
antibody-detection assays, on sera, plasma, urine, saliva etc., they are
highly sensitive and specific. The specification also describes 69
previously unrecognised immunogenic cluster families. H. pylori antigens
are used to detect H. pylori-specific antibodies, for diagnosing
infection or to confirm eradication of infection, and in vaccines to
protect against H. pylori infection and related diseases (gastritis,
peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ Sequence 188 AA;

Query Match      22.5%; Score 181; DB 20; Length 188;
Best Local Similarity 28.0%; Pred. No. 9.8e-14;
Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;

```

```

QY 13 LIEPFSEE-----SLOPAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
DB 18 mispfcekgvqknvisyglssygydirvgsefmldknalidpknfdpnnatkidaske 77
QY 48 GKVVIPPREYALITLERIKLPDDVMDGMKIRSSLAREGVI--GSFAWVDPGWDGNLTMLY 107
DB 78 gyfilpanafalantleyfknmpkdtlaiclgkstyarcgiiinvtpfepfegyitais 137
QY 108 NASNEPVELRYGERFVQIATRLREGPARNPYR---GNYQGSTRLAESK 152
DB 138 nttnlpakvyaneaglaqvfflqgdemceqsykdrqgkygvgvgitlpk 185

RESULT 6
AAW89897
ID AAW89897 standard; Protein: 190 AA.
XX
AC AAW89897;
XX
XX
DT 18-FEB-1999 (first entry)
DE Antigen 3 from cluster 28c.
XX
KW Antigen: immunogenic cluster family; vaccine; gastritis; diagnosis;
KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.
XX
OS Helicobacter pylori.
XX
PN W09849314-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98WO-US08487.
XX
PR 14-OCT-1997; 97US-0061958.
XX
PR 25-APR-1997; 97US-0045107.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Chow TP, Fry KE, Lim MY, McAtee CP;
XX
DR WPI; 1999-009433/01.
XX
XX
New Helicobacter pylori antigens and related nucleic acid sequences
- useful in serological diagnosis and protective vaccines, providing
long-lasting immune response
Claim 1; Page 236-237; 402pp; English.
XX
XX
The present sequence represents a Helicobacter pylori antigenic protein
that is characterised by immunoreactivity with H. pylori-positive
antisera. The proteins are highly immunogenic and induce a long-lasting
immune response that persists even after antimicrobial treatment. In
antibody-detection assays, on sera, plasma, urine, saliva etc., they are
highly sensitive and specific. The specification also describes 69
previously unrecognised immunogenic cluster families. H. pylori antigens
are used to detect H. pylori-specific antibodies, for diagnosing
infection or to confirm eradication of infection, and in vaccines to
protect against H. pylori infection and related diseases (gastritis,
peptic ulcer, gastric adenocarcinoma/lymphoma).
XX
SQ Sequence 190 AA;

Query Match      22.5%; Score 181; DB 20; Length 190;
Best Local Similarity 28.0%; Pred. No. 9.9e-14;
Matches 47; Conservative 29; Mismatches 64; Indels 28; Gaps 4;

QY 13 LIEPFSEE-----SLOPAGYDLRVGRE-----AFVKGK-----LIDVEKE 47
DB 20 mispfcekgvqknvisyglssygydirvgsefmldknalidpknfdpnnatkidaske 79
QY 48 GKVVIPPREYALITLERIKLPDDVMDGMKIRSSLAREGVI--GSFAWVDPGWDGNLTMLY 107

```

```

Db      80 gyfilpanafalatieyfkmpkdtlaiclgkstyarcgiiinvtpfepegvities 139
QY      108 NASNEPVELRYGERVOIAFIRLEGPARNPYR--GNYQGSTRLAFSK 152
Db      140 ntnlpakvyaneqiaqvflqdemceqsykdrqgkyqgvgitlpx 187

RESULT  7
ID      AAY37187 standard; Protein; 190 AA.
AC      AAY37187;
XX
XX
DT      07-OCT-1999 (first entry)
DE
DE
DE
XX
XX
XX
KW      Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW      paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
KW      nongonococcal urethritis; epidymitis; cervicitis; salpingitis;
KW      bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS      Chlamydia trachomatis.
XX
XX
PN      WO9928475-A2.
XX
XX
PD      10-JUN-1999.
XX
XX
PF      27-NOV-1998; 98WO-IB01939.
XX
XX
PR      04-NOV-1998; 98US-0107077.
PR      28-NOV-1997; 97FR-0015041.
PR      17-DEC-1997; 97FR-0016034.
XX
XX
PA      (GEST ) GENSET.
XX
XX
PI      Griffais R;
XX
XX
WPI; 1999-371125/31.
XX
XX
Genome sequence of Chlamydia trachomatis
XX
XX
Disclosure; Page 961; 1755pp; English.
XX
XX
AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
can also be used to control growth of the microorganism. Chlamydia
trachomatis is responsible for a large number of diseases, e.g. eye
diseases such as conventional trachoma, nonendemic trachoma,
paratrachoma, and inclusion conjunctivitis; genital diseases such as
nongonococcal urethritis, epidymitis, cervicitis, salpingitis,
perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;
and venereal lymphogranulomatosis. The polypeptides of the invention
may be of use in treating these diseases.
XX
XX
Sequence 190 AA;

Query Match 18.5%; Score 149.5; DB 20; Length 190;
Best Local Similarity 26.1%; Pred. No. 6.5e-10;
Matches 46; Conservative 31; Mismatches 60; Indels 39; Gaps 6;

QY      5 DWKIRK-----EILIEPSEE-----SLQAGYDLRVGRE----- 34
Db      7 nw-irkmaieegmlepfadsqvkllhpetgeklisgydlrlsrefkvtfnvynsl 65
QY      35 ----AFVKGKLIIVEREGKVIPREVALILTLERIKLPDDVMGDMKIRSSLAREGVIGS 90
Db      66 vdpkctfedalisi-vddvcilpnsfalarsveyfrprnvlvtmcigkstyarcgiiinv 124
QY      91 FAWVDPGWDGNLTMLYNASNEPVELRYGERVOIAFIRLEGPARNPY---RGNYQ 143

```

```

Db      125 vtpfepegvvtieisnttplpakvyaneqiaqvlfegdaacdvsyaerqgkyq 180

RESULT  8
ID      AAY34994 standard; Protein; 190 AA.
XX
XX
AC      AAY34994;
XX
XX
DT      13-SEP-1999 (first entry)
DE
DE
DE
XX
XX
XX
KW      Protein involved in intermediate metabolism of nucleotides.
KW      Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW      sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW      vaccine; neutralising epitope.
XX
OS      Chlamydia pneumoniae.
XX
XX
PN      WO9927105-A2.
XX
XX
PD      03-JUN-1999.
XX
XX
PF      20-NOV-1998; 98WO-IB01890.
XX
XX
PR      04-NOV-1998; 98US-0107078.
PR      21-NOV-1997; 97FR-0014673.
XX
XX
PA      (GEST ) GENSET.
XX
XX
PI      Griffais R;
XX
XX
WPI; 1999-357842/30.
XX
XX
Genome sequence of Chlamydia pneumoniae
XX
XX
Page 901; Disclosure; 1912pp; English.
XX
XX
AAY34584-Y35879 represent the proteins encoded by all the open reading
frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
C. pneumoniae causes respiratory disease such as pneumonia and
bronchitis and is thought to be a contributing factor in heart
disease, sarcoidosis, sinusitis, purulent otitis media, erythema
nodosum or pharyngitis. The polypeptides encoded by the open reading
frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
immunogenic compositions as vaccines. Vectors containing C. pneumoniae
nucleotide sequences can also be used as immunogenic compositions,
especially where the vector directs the expression of a neutralising
epitope of C. pneumoniae.
XX
XX
Sequence 190 AA;

Query Match 17.1%; Score 138; DB 20; Length 190;
Best Local Similarity 24.4%; Pred. No. 1.6e-08;
Matches 40; Conservative 31; Mismatches 59; Indels 34; Gaps 5;

QY      13 LIEPFF-----SEES-----LQAGYDLRVGRE-----AFVKGKLI 42
Db      18 mihpfvngvgvnnveetgeklisgydlrlsrefkvtfnvynsvvdpkctedifi 77
QY      43 DVEKEGKVITPREYALILTLERIKLPDDVMGDMKIRSSLAREGVIGSFAWVDPGWDGNL 102
Db      78 si-tddvcilpnsfalarsveyfrprnvlvtmcigkstyarcgiiinvtpfepegv 136
QY      103 TMLYNASNEPVELRYGERVOIAFIRLEGPARNPY---RGNYQ 143
Db      137 tieisnttplakiyaneqiaqvlfessttcevsyadrkqkyq 180

RESULT  9
AAY29169

```


XX DE Amino acid sequence of a virulence factor encoded by ORF17793c.
 XX KW Human pathogen; virulence polypeptide; virulence factor;
 XX KW pathogenic infection; Pseudomonas aeruginosa infection.
 XX OS Pseudomonas aeruginosa.
 XX PN WO9927129-A1.
 XX PD 03-JUN-1999.
 XX PF 25-NOV-1998; 98WO-US25247.
 XX PR 25-NOV-1997; 97US-0066517.
 XX PA (GEO) GEN HOSPITAL CORP.
 XX PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
 XX PI Rahme LG, Tan M, Tsongalis J;
 XX DR WPI; 1999-357851/30.
 XX PF Virulence factors useful in developing disease treatments
 XX PS Disclosure; Fig 4; 228pp; English.
 XX CC The present sequence represents a Pseudomonas aeruginosa polypeptide
 CC CC sequence. P. aeruginosa is an opportunistic human pathogen present in
 CC CC soil water and plants. The specification describes virulence polypeptides
 CC CC and nucleic acid sequence encoding such polypeptides. These sequences
 CC CC can be used to identify a compound which is capable of decreasing the
 CC CC expression of a pathogenic virulence factor. Compounds that inhibit
 CC CC the expression or activity of virulence factor polypeptides can be
 CC CC used to treat pathogenic infections, especially where the infection
 CC CC is a P. aeruginosa infection.
 CC CC note: the sequences given in the specification were poorly legible, and
 CC CC in some instances assumptions were made as to the identity of the
 CC CC residue; it is therefore possible that the sequence given below is
 CC CC not entirely correct.
 XX SQ Sequence 181 AA;
 Query Match 13.9%; Score 112; DB 20; Length 181;
 Best Local Similarity 24.9%; Pred. No. 2.1e-05;
 Matches 43; Conservative 28; Mismatches 76; Indels 26; Gaps 5;
 QY 2 LLPDKIRKEILIEPFS-----EESLPAGYDLRVGREA-----FVKGLIDVEK 46
 DB 13 lvaqcalcektlfdelskiglhtyepkspasasivgyshpaptffilepk-----ei 68
 QY 47 EGVVIPPPEVALITLERIKLPDDVGMKIRSLAR--EGVIGSFAWDPGWDGNLTL 104
 DB 69 qgnlvksgeqvitsckhrkypidylgvlqtkglarlfvqtcndgqvepgfdgyvtl 128
 QY 105 MLYNASNPVELRYGERVQIAFIRLEGPARNVGNV-----OGSTRLEFSK 152
 DB 129 eivmnpwtieipavsdiaqlvkcstasepyngrymdaakkgptlavirk 181
 RESULT 12
 AAR70144
 ID AAR70144 standard; Protein; 141 AA.
 XX AC AAR70144;
 XX DT 23-SEP-1995 (first entry)
 XX DE Human dUTPase protomer.
 XX KW Deoxyuridine-triphosphate-nucleotidohydrolase; dUTPase;
 KW antimicrobial; anticancer; cytostatic; screening.

XX OS Homo sapiens.
 XX PN CA2126001-A.
 XX PD 28-JAN-1995.
 XX PF 16-JUN-1994; 94CA-2126001.
 XX PR 27-JUL-1993; 93US-0097220.
 XX PA (ALLX) ALLELIX BIOPHARMACEUTICALS INC.
 XX PI Climie S, Vandenbergh E;
 XX DR WPI; 1995-115928/16.
 XX DR N-PSDB; AAR83219.
 XX PF Pure native human deoxyuridine triphosphate nucleotidohydrolase
 PF - used to identify inhibitors, potentially useful as
 PF antimicrobial and anticancer agents
 XX PS Disclosure; Fig.1; 18pp; English.
 XX CC DNA encoding human dUTPase (given in AAR83219) was cloned into pUC19
 CC and recombinants used to transform E. coli BW 286. A transformant
 CC selected for growth at 39 deg was cultured overnight. Recombinant
 CC human dUTPase was purified and shown to be identical in sequence to
 CC the natural protomeric peptide (AAR70144).
 XX SQ Sequence 141 AA;
 Query Match 13.8%; Score 111; DB 16; Length 141;
 Best Local Similarity 30.7%; Pred. No. 1.9e-05;
 Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;
 QY 21 SLQAGYDLRVGREA-FVKGLIDVEKGVVIPPPEVALITLERIKLPDDVGMKIRSR 80
 DB 19 saraagydllysayd-----yttppmekavvktldialpscygrvapr 63
 QY 81 SLAREGVIGSFAWV-DPGWDGNLTLMLYNASNPVELRYGERVQIAFIRLEGP 133
 DB 64 glaakhfidvgagvldedyrngvgnvlfngfgekfevkkgdriaqlicerifyp 117
 RESULT 13
 AAW30280
 ID AAW30280 standard; Protein; 164 AA.
 XX AC AAW30280;
 XX DT 14-APR-1998 (first entry)
 XX DE Human dUTPase (nuclear form);
 KW dUTPase; DUT-N; human; cell proliferation; inhibitor; cytostatic;
 KW antiviral; antibacterial; antifungal; antibiotic.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Modified-site 11
 FT Peptide /note= "O-phosphorylated"
 FT 1.15
 FT /note= "N-terminal peptide unique to DUT-N"
 XX PN WO9736916-A1.
 XX PD 09-OCT-1997.
 XX PF 26-MAR-1997; 97WO-US04886.

PR 26-MAR-1997; 97US-0824405.
 XX 29-MAR-1996; 96US-0014748.
 PA (UYNE-) UNIV NEW JERSEY.
 XX Caradonna SJ, Ladner RD, Lynch F;
 XX WPI; 1997-503040/46.
 DR N-PSDB; AAT90987.
 XX Nucleic acid encoding two isoform(s) of human dUTPase - used in
 PT assays for assessing status of cell proliferation, effect of
 PT anticancer agents and progress of treatment
 XX
 PS Claim 6; Fig 1A-B; 89pp; English.
 XX This polypeptide comprises the nuclear form, designated DUT-N, of
 CC human deoxyuridine triphosphate nucleotidohydrolase (dUTPase). Its
 CC amino acid sequence was deduced from a T-cell cDNA clone (see
 CC AAT90987). A single gene codes for 2 isoforms of human dUTPase, the
 CC nuclear form (DUT-N) and a cytoplasmic form (DUT-M, see AAW30281)
 CC which is targeted to the mitochondria. The isoforms arise by the
 CC use of alternative 5' exons which then join in a common sequence.
 CC The DUT-N isoform is produced at low or undetectable levels during
 CC non-proliferative periods of the cell cycle, and can be used as a
 CC marker for cellular proliferation. It is phosphorylated by p34cdc2
 CC at a Ser residue within the consensus sequence for cyclin-dependent
 CC kinase phosphorylation. Measuring the amount of dUTPase in a cell
 CC can be used to determine its proliferative status (or changes in
 CC this status), particularly in nutrient-deprived or neoplastic cells
 CC (claimed). This method can be used for determining the effect of
 CC anticancer agents (A) on cells, or monitoring the effect of (A)
 CC (claimed). Agents that inhibit dUTPase are potential
 CC antineoplastic, antiviral, antibacterial and antifungal agents.
 CC The dUTPase can also be used to raise antibodies for assaying the
 CC enzyme and for drug screening.
 XX Sequence 164 AA;
 SQ

Query Match 13.8%; Score 111; DB 18; Length 164;
 Best Local Similarity 30.7%; Pred. No. 2.4e-05;
 Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;
 Qy 21 SLQPGYDLRVGREAFVKGKLDIDVEKEGVVIPPVEYALITLERIKLPDDVMGDMKIRS 80
 Db 42 saraagdylysdyd-----ytipmekavvktidialpsegcygrvapr 86
 Qy 81 SLAREGVIGSAFW-DPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
 Db 87 glaakhfidvgagviedyrgnvgvifnfgkfevkkgdriaqlcerifyp 140

RESULT 14
 AAW30281
 ID AAW30281 standard; Protein; 252 AA.
 AC AAW30281;
 XX 14-APR-1998 (first entry)
 XX Human dUTPase (mitochondrial form).
 XX dUTPase; DUT-M; human; cell proliferation; inhibitor; cytostatic;
 KW antiviral; antibacterial; antifungal; antibiotic.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 PH 70...93 "DUT-M unique sequence 5' to common Glu
 FT /note- residue of DUT-M and DUT-N"
 FT
 XX

PN W09736916-A1.
 XX 09-OCT-1997.
 XX 26-MAR-1997; 97WO-US04886.
 XX 26-MAR-1997; 97US-0824405.
 PR 29-MAR-1996; 96US-0014748.
 XX (UYNE-) UNIV NEW JERSEY.
 XX Caradonna SJ, Ladner RD, Lynch F;
 XX WPI; 1997-503040/46.
 DR N-PSDB; AAT90989.
 XX Nucleic acid encoding two isoform(s) of human dUTPase - used in
 PT assays for assessing status of cell proliferation, effect of
 PT anticancer agents and progress of treatment
 XX
 PS Claim 6; Fig 2A-B; 89pp; English.
 XX This polypeptide comprises the mitochondrial form, designated
 CC DUT-M, of human deoxyuridine triphosphate nucleotidohydrolase
 CC (dUTPase). Its amino acid sequence was deduced from a fibroblast
 CC cDNA clone (see AAT90989). A single gene codes for 2 isoforms of
 CC human dUTPase, the nuclear form (DUT-N, see AAW30280) and the
 CC cytoplasmic form (DUT-M) which is targeted to the mitochondria.
 CC The isoforms arise by the use of alternative 5' exons which then
 CC join in a common sequence. The DUT-M isoform is detectable in
 CC proliferating and non-proliferating cells, but is produced in
 CC increased amounts during periods of non-proliferation in certain
 CC tumours. Unlike DUT-N, it is not phosphorylated in vivo.
 CC Measuring the amount of dUTPase in a cell can be used to determine
 CC its proliferative status (or changes in this status), particularly
 CC in nutrient-deprived or neoplastic cells (claimed). This method
 CC can be used for determining the effect of anticancer agents (A) on
 CC cells, or monitoring the effect of (A) (claimed). Agents that
 CC inhibit dUTPase are potential antineoplastic, antiviral,
 CC antibacterial and antifungal agents. The dUTPase can also be used
 CC to raise antibodies for assaying the enzyme and for drug screening.
 XX Sequence 252 AA;
 SQ

Query Match 13.8%; Score 111; DB 18; Length 252;
 Best Local Similarity 30.7%; Pred. No. 4.5e-05;
 Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;
 Qy 21 SLQPGYDLRVGREAFVKGKLDIDVEKEGVVIPPVEYALITLERIKLPDDVMGDMKIRS 80
 Db 130 saraagdylysdyd-----ytipmekavvktidialpsegcygrvapr 174
 Qy 81 SLAREGVIGSAFW-DPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
 Db 175 glaakhfidvgagviedyrgnvgvifnfgkfevkkgdriaqlcerifyp 228

RESULT 15
 AAW71667
 ID AAW71667 standard; Protein; 1105 AA.
 AC AAW71667;
 XX 26-APR-1991 (first entry)
 XX Translation product of pol transcript of VISNA sheep lentivirus.
 DE VISNA; vaccine; reverse transcriptase; Ovis aries; lentivirus;
 KW meningoencephalitis of sheep.
 XX visna lentivirus.
 OS

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:07:51 ; Search time 19.24 Seconds
(without alignments)
182.459 Million cell updates/sec

Title: US-08-957-709-71
Perfect score: 806
Sequence: 1 MLLPDWKIRKEILIEFSEE.....PYRGNYQGSTRLAFSKRKKL 156

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	533	66.1	246	4	US-08-822-774-44
2	160	19.9	42	4	US-08-822-774-41
3	119	14.8	27	4	US-08-822-774-38
4	114.5	14.2	138	2	US-08-824-405-11
5	111	13.8	138	2	US-08-824-405-10
6	111	13.8	164	2	US-08-824-405-2
7	111	13.8	232	2	US-08-824-405-4
8	87.5	10.9	1150	4	US-09-238-303-9
9	87	10.8	147	2	US-08-824-405-12
10	81.5	10.1	144	2	US-08-824-405-13
11	79.5	9.9	147	2	US-08-824-405-6
12	78.5	9.7	142	1	US-08-307-499-22
13	78.5	9.7	142	4	US-09-299-268-22
14	76.5	9.5	14	4	US-08-822-774-54
15	76.5	9.5	659	1	US-08-258-639A-4
16	76.5	9.5	659	2	US-08-900-951-4
17	76.5	9.5	659	5	PCT-US95-07391A-4
18	74	9.2	17	4	US-08-822-774-11
19	73.5	9.1	151	2	US-08-824-405-7
20	72	8.9	424	1	US-08-453-472-8
21	72	8.9	424	1	US-08-038-948-2
22	72	8.9	424	1	US-08-038-948-5
23	72	8.9	424	1	US-08-453-952-8
24	72	8.9	424	2	US-08-862-903-8
25	71	8.8	141	2	US-08-824-405-8
26	71	8.8	142	2	US-08-824-405-9
27	71	8.8	424	1	US-08-453-472-7

28	71	8.8	424	1	US-08-038-948-4	Sequence 4, Appli
29	71	8.8	424	1	US-08-453-952-7	Sequence 7, Appli
30	71	8.8	424	2	US-08-862-903-7	Sequence 7, Appli
31	71	8.8	424	2	US-08-484-158B-61	Sequence 61, Appli
32	70.5	8.7	415	2	US-08-484-993B-8	Sequence 8, Appli
33	70.5	8.7	415	2	US-08-484-158B-8	Sequence 8, Appli
34	70.5	8.7	415	2	US-08-484-596A-8	Sequence 8, Appli
35	70.5	8.7	415	2	US-08-480-150A-8	Sequence 8, Appli
36	70.5	8.7	415	3	US-08-458-731-8	Sequence 8, Appli
37	70.5	8.7	415	3	US-08-149-223A-8	Sequence 8, Appli
38	70	8.7	820	2	US-08-374-483-2	Sequence 2, Appli
39	70	8.7	820	2	US-08-374-483-5	Sequence 5, Appli
40	69	8.6	414	2	US-08-599-171A-25	Sequence 25, Appli
41	69	8.6	414	2	US-08-646-590B-25	Sequence 25, Appli
42	69	8.6	414	3	US-09-069-226-25	Sequence 25, Appli
43	69	8.6	414	4	US-09-412-184-25	Sequence 2, Appli
44	68.5	8.5	553	2	US-08-586-272-2	Sequence 2, Appli
45	68.5	8.5	553	3	US-09-082-969-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-822-774-44
Sequence 44, Application US/08822774
Patent No. 6183997
GENERAL INFORMATION:
APPLICANT: HOGREPE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
TITLE OF INVENTION: and Methods for Purifying and Identifying Same
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
ADDRESSEE: Lenahan, P.L.L.C.
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ. ID. NO. 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 246 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: YES
US-08-822-774-44

Query Match 66.1%; Score 533; DB 4; Length 246;
Best Local Similarity 98.1%; Pred. No. 9.5e-59;
Matches 102; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 53 PPREYALILTLERIKLPDDYGMGMKIRSLAREGVGSFAWDFGWDGNTLMLYNASNE 112
Db 1 PPREYALILTLERIKLPNNVGMGMKIRSLAREGVGSFAWDFGWDGNTLMLYNASNE 60

```
Oy 113 PVELRYGERFVOIAFIRLEGPARPNRYGNYGSTRFLAFSKRKL 156
      |||||||
Db 61 PVELRYGERFVOIAFIRLEGPARPNRYGNYGSTRFLAFSKRKL 104

RESULT 2
US-08-822-774-41 Application US/08822774
; Sequence 41, Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESS: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-822-774-41

Query Match 19.9%; Score 160; DB 4; Length 42;
Best Local Similarity 96.9%; Pred. No. 3.9e-13;
Matches 31; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MLLPDWKIRKEILIEPFSEESLPAGYDLRVG 32
      |||||||
Db 11 MLLPDWKIRKEILIEPFSEEWLPAGYDLRVG 42

RESULT 3
US-08-822-774-38 Application US/08822774
; Sequence 38, Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESS: Lenahan, P.L.L.C.
```

```
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHEICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-822-774-38

Query Match 14.8%; Score 119; DB 4; Length 27;
Best Local Similarity 88.3%; Pred. No. 2.8e-08;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 30 RVGREAFVKGKLDVKEGKVIPPPE 56
      |||||||
Db 1 RVGREAXVKGKLLIEVEGKVXIPPPE 27

RESULT 4
US-08-824-405-11 Application US/08824405
; Sequence 11, Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/824,405
FILING DATE: 26-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
```

```
;
;
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-824-405-11

Query Match 14.2%; Score 114.5; DB 2; Length 138;
Best Local Similarity 22.8%; Pred. No. 1.1e-06;
Matches 28; Conservative 34; Mismatches 46; Indels 15; Gaps 1;

QY 19 EESLQAGYDLRVGREAFVKGKLDVKEGKVIPPPEYALILTLEIKLPDDVMGDMKI 78
Db 14 EKRDAGEDLCVPYD-----IMPVSDTKIIPTDVKIQVPPNSFGWVTG 58

QY 79 RSSLAREGVIGSFANWDPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY 138
Db 59 KSSMAKOGLLINGIIDEGYGTGCTNIGKSNIKLIEGQKFAQLIILQHHNSRQPW 118

QY 139 RGN 141
Db 119 DEN 121

RESULT 5
US-08-824-405-10
; Sequence 10, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids

;
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-824-405-10

Query Match 13.8%; Score 111; DB 2; Length 138;
Best Local Similarity 24.8%; Pred. No. 3.1e-06;
Matches 32; Conservative 24; Mismatches 53; Indels 20; Gaps 2;

QY 19 EESLQAGYDLRVGREAFVKGKLDVKEGKVIPPPEYALILTLEIKLPDDVMGDMKI 78
Db 14 OKRAEDAGYDLICPQE-----ISIPAGQVKRIADLKLKQDQWAMIGT 58

QY 79 RSSLAREGVIGSFANWDPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGPARNPY 138
Db 59 KSFANKGVFVGGIIDSIGYGTGIIQVYVYNSNKNKEVVIPOGRKFAQLILMPLTHEELEPW 118

QY 139 RGNVQGST 147
Db 119 -----GETR 122

RESULT 6
US-08-824-405-2
; Sequence 2, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-824-405-2

Query Match 13.8%; Score 111; DB 2; Length 164;
Best Local Similarity 30.7%; Pred. No. 4e-06;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;
```

```
QY 21 SLQAGYDLRVGREAFVKGKLDIVKEGKVVIPPREYALITLTERIKLPDDVGMKIRS 80
Db 42 SARAAGVDLYSAYD-----YTIPPMKAVVKTDIQAIPSGCYGRVAPRS 86
QY 81 SLAREGVIGSFAMW-DPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 87 GLAAKHFDVAGVIDEDYRGNGVWLFNFGKEKFEVKKGDRIAQLICERIFYP 140

RESULT 7
US-08-824-405-4
; Sequence 4, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; OTHER INFORMATION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-824-405-4

Query Match 13.8%; Score 111; DB 2; Length 252;
Best Local Similarity 30.7%; Pred. No. 7.6e-06;
Matches 35; Conservative 17; Mismatches 46; Indels 16; Gaps 2;

QY 21 SLQAGYDLRVGREAFVKGKLDIVKEGKVVIPPREYALITLTERIKLPDDVGMKIRS 80
Db 130 SARAAGVDLYSAYD-----YTIPPMKAVVKTDIQAIPSGCYGRVAPRS 174
QY 81 SLAREGVIGSFAMW-DPGWDGNLTMLYNASNEPVELRYGERFVQIAFIRLEGP 133
Db 175 GLAAKHFDVAGVIDEDYRGNGVWLFNFGKEKFEVKKGDRIAQLICERIFYP 228

RESULT 8
US-09-238-303-9
; Sequence 9, Application US/09238303B
```

```
; Patent No. 6284253
; GENERAL INFORMATION:
; APPLICANT: Bair, Margaret C.
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequenc
; FILE REFERENCE: 18617.0059
; CURRENT APPLICATION NUMBER: US/09/238,303B
; CURRENT FILING DATE: 1999-01-28
; EARLIER APPLICATION NUMBER: US 60/072,927
; EARLIER FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 9
; LENGTH: 1150
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: protein encoded by the pol gene of a recombinant viral
; OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat felin
; OTHER INFORMATION: immunodeficiency virus
; US-09-238-303-9

Query Match 10.9%; Score 87.5; DB 4; Length 1150;
Best Local Similarity 24.1%; Pred. No. 0.062;
Matches 26; Conservative 23; Mismatches 40; Indels 19; Gaps 3;

QY 23 QPAGYDLRVGREAFVKGKLDIVKEGKVVIPPREYALITLTERIKLPDDVGMKIRSSL 82
Db 750 EDAGYDILLA-----EQDIHLMPEGEVRIVPTGVRLMLPKGHGWMVVGKSSI 794
QY 83 AREG--VIGSAFWDPGWDGNLTMLYNASNEPVELRYGERFVQIAET 128
Db 795 AKOGLDVLGGV--IDEGYRGEIGVIMINLQKRSTLKEKQKVAQLIII 840

RESULT 9
US-08-824-405-12
; Sequence 12, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; OTHER INFORMATION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
```


US-08-307-499-22

```

; Sequence 22, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,499
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 142 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-307-499-22

```

```

Query Match          9.7%; Score 78.5; DB 1; Length 142;
Best Local Similarity 25.2%; Pred. No. 0.038;
Matches 33; Conservative 14; Mismatches 49; Indels 35; Gaps 4;

QY 25 AGDYLVRGFEAFVKGLIDVEKEGVVIPPPEYALILTLERIKLPDDVMDGDKIRSSLAR 84
Db 25 AGDYLVSAYSVTVK-----PYNRLIVRTDCLMIPDKCYGRISPRSGLSL 69

QY 85 E-----GVIGSFAWDPGWDGNLTMLYNASNEPVRLYGERFVQIAFIRLEGP---- 133
Db 70 NYNIDIGGGVI-----DSDYRGEIGIVFINCGSDFNKVGDRIAQIIIFERVEYPIIMEE 123

QY 134 ---ARNPYRGN 141
Db 124 VKCLEDTERGN 134

RESULT 13
US-09-299-268-22
; Sequence 22, Application US/09299268

```

```

; Patent No. 6217882
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi uela, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,268
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/901,127
; FILING DATE:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 142 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-268-22

```

```

Query Match          9.7%; Score 78.5; DB 4; Length 142;
Best Local Similarity 25.2%; Pred. No. 0.038;
Matches 33; Conservative 14; Mismatches 49; Indels 35; Gaps 4;

QY 25 AGDYLVRGFEAFVKGLIDVEKEGVVIPPPEYALILTLERIKLPDDVMDGDKIRSSLAR 84
Db 25 AGDYLVSAYSVTVK-----PYNRLIVRTDCLMIPDKCYGRISPRSGLSL 69

QY 85 E-----GVIGSFAWDPGWDGNLTMLYNASNEPVRLYGERFVQIAFIRLEGP---- 133
Db 70 NYNIDIGGGVI-----DSDYRGEIGIVFINCGSDFNKVGDRIAQIIIFERVEYPIIMEE 123

QY 134 ---ARNPYRGN 141
Db 124 VKCLEDTERGN 134

RESULT 14
US-08-822-774-54
; Sequence 54, Application US/08822774

```


; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; US-08-822-774-54

Query Match 9.5%; Score 76.5; DB 4; Length 14;
Best Local Similarity 93.3%; Pred. No. 0.0023;
Matches 14; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 91 FAWVDPGWDGNLTLM 105
Db 1 FAWVDPGWDGN-TLM 14

RESULT 15
US-08-258-639A-4
; Sequence 4, Application US/08258639A
; Patent No. 5681733
; GENERAL INFORMATION:
; APPLICANT: Su, Hongsheng
; APPLICANT: Blain, Francoise
; APPLICANT: Bennett, Clark
; APPLICANT: Gu, Kangfu
; APPLICANT: Zimmermann, Joseph
; APPLICANT: Musil, Roy
; TITLE OF INVENTION: Nucleic Acid Sequences And Expression
; TITLE OF INVENTION: Systems For Heparinase II And Heparinase III Derived From
; TITLE OF INVENTION: Flavobacterium heparinum
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hale and Dorr
; STREET: 1455 Pennsylvania Avenue, N.W.
; CITY: Washington, D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,639A
; FILING DATE: 10 JUNE 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Healey, William J.
; REGISTRATION NUMBER: 36,160
; REFERENCE/DOCKET NUMBER: 104385.116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)942-8400
; TELEFAX: (202)942-8484
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 659 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-258-639A-4

Query Match 9.5%; Score 76.5; DB 1; Length 659;
Best Local Similarity 23.2%; Pred. No. 0.66;
Matches 38; Conservative 19; Mismatches 42; Indels 65; Gaps 8;

QY 4 PDWK-----IRKEILIEPFSEESLOP-----AGYDLRVGREAFVKGLIDVEKEGK 49
Db 263 PRWRQTGISVLNTEIKKQVYADGMQFE-LSPIYHVAIDIFL--KAYGSAKRVNLEKEF- 318
QY 50 VVIPPREYA-----LITLERIKLPD---DVMGDMKIRSSLAREGVIGSPAWVDP----- 96
Db 319 -----POSYVQTVENMIMALISISLPDYNTPFMGDSMTDKNFRMAQFASWARYFPANQAI 374
QY 97 -----GWDGNLTLMLYNAS 110
Db 375 KYFATDGKQKAPNFLSKALSNAGFYTFRSGWKNATVMVLKAS 418

Search completed: January 31, 2002, 13:07:52
Job time: 44 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:15 ; Search time 46.78 seconds
(without alignments)
10.189 Million cell updates/sec

Title: US-08-957-709-72
Perfect score: 13
Sequence: 1 XGXXDXXGXXXX 13

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	2	15.4	34	1 PYSB_METBA	P80522 methanosarc
2	2	15.4	40	1 RIPI_SAPOF	P98185 saponaria o
3	2	15.4	466	1 RBL_CORKO	Q33369 cornus kous
4	1	7.7	3	1 GRWM_HUMAN	P01157 homo sapien
5	1	7.7	3	1 LUXE_VIBFI	P24272 vibrio fisc
6	1	7.7	4	1 ACHI_ACHFU	P35904 achatina fu
7	1	7.7	4	1 DCML_PSECH	P19916 pseudomonas
8	1	7.7	4	1 EOSI_HUMAN	P02731 homo sapien
9	1	7.7	5	1 AL14_CARMA	P81817 carcinus ma
10	1	7.7	5	1 BIOA_CITFR	P13071 citrobacter
11	1	7.7	5	1 BIOA_SALTY	P12677 salmonella
12	1	7.7	5	1 PAP2_PARMA	P81864 pardachirus
13	1	7.7	5	1 TPIS_CANFA	P54714 canis fami
14	1	7.7	5	1 TRM3_ECOLI	P13973 escherichia
15	1	7.7	5	1 UF01_MOUSE	P38639 mus musculu
16	1	7.7	5	1 UXA4_CHLTR	P38005 chlamydia t
17	1	7.7	6	1 ASP2_LACSN	P82655 lactobacill
18	1	7.7	6	1 CIP1_MYTED	P13736 mytilus edu
19	1	7.7	6	1 CIP2_MYTED	P13737 mytilus edu
20	1	7.7	6	1 FARP_MONEX	P41966 montezia ex
21	1	7.7	6	1 LOK1_LOCOMI	P41491 locusta mig
22	1	7.7	6	1 TRP1_PSEPU	P36414 pseudomonas
23	1	7.7	7	1 ALL2_CARMA	P81805 carcinus ma
24	1	7.7	7	1 ALL3_CARMA	P81806 carcinus ma
25	1	7.7	7	1 ALL4_CARMA	P81807 carcinus ma
26	1	7.7	7	1 ALL5_CARMA	P81808 carcinus ma
27	1	7.7	7	1 ALL7_CVDPO	P82158 cydia pomon
28	1	7.7	7	1 CHOX_ALCSP	P16101 alcaligenes
29	1	7.7	7	1 FAR1_HELTI	P41871 helisoma tr
30	1	7.7	7	1 FAR2_PROCL	P38498 procambur
31	1	7.7	7	1 FAR5_HIRME	P42564 hirudo medi
32	1	7.7	7	1 FARP_CALVO	P41866 calliphora
33	1	7.7	7	1 IGAO_DACDE	P06294 dactylium d

34	1	7.7	7	1 LANC_CARUI	P36960 carnobacter
35	1	7.7	7	1 MNPI_LEPDE	P42984 leptinotars
36	1	7.7	7	1 UF03_MOUSE	P38641 mus musculu
37	1	7.7	7	1 UF04_MOUSE	P38642 mus musculu
38	1	7.7	7	1 UH11_RAT	P56576 rattus norv
39	1	7.7	7	1 UN06_PINPS	P81675 pinus pinas
40	1	7.7	8	1 ACI_THUAL	P18691 thunnus alb
41	1	7.7	8	1 ACT_CARMA	P80709 carcinus ma
42	1	7.7	8	1 AKHG_GRYBI	P14086 gryllus bim
43	1	7.7	8	1 AKH_MELML	P25423 melolontha
44	1	7.7	8	1 AKH_TABAT	P14595 tabanus atr
45	1	7.7	8	1 AL12_CARMA	P91815 carcinus ma

ALIGNMENTS

RESULT 1					
PYSB_METBA					
ID	PYSB_METBA	STANDARD;	PRT;	34 AA.	
AC	P80522;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	PYRUVATE SYNTHASE BETA CHAIN (EC 1.2.7.1) (FRAGMENT).				
OS	Methanosarcina barkeri.				
OC	Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;				
OC	Methanosarcina.				
OX	NCBI_TaxID=2208;				
RN	[1]				
RP	SEQUENCE.				
RC	STRAIN=FUSARO / DSM 804;				
RX	MEDLINE=96203906; PubMed=8620891;				
RA	Bock A.-K., Kunow J., Glasemacher J., Schoenheit P.;				
RT	*catalytic properties, molecular composition and sequence alignments				
RT	of pyruvate: ferredoxin oxidoreductase from the methanogenic archaeon				
RT	Methanosarcina barkeri (strain Fusaro).";				
RL	Eur. J. Biochem. 237:35-44(1996).				
CC	-I- CATALYTIC ACTIVITY: PYRUVATE + COA + OXIDIZED FERREDOXIN =				
CC	ACETYL-COA + CO(2) + REDUCED FERREDOXIN.				
CC	-I- SUBUNIT: HETEROTETRAMER OF ONE ALPHA, ONE BETA, ONE DELTA AND				
CC	ONE GAMMA CHAIN.				
CC	-I- MISCELLANEOUS: IT ALSO CATALYZES THE OXIDATION OF 2-OXOBUTYRATE.				
KW	Oxidoreductase.				
FT	NON_TER 34 34				
SQ	SEQUENCE 34 AA; 3372 MW; C913DA5DDELD75FE CRC64;				

Query Match 15.4%; Score 2; DB 1; Length 34;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2 GXXD 5
Db	18 GXXD 21

RESULT 2					
RIPI_SAPOF					
ID	RIPI_SAPOF	STANDARD;	PRT;	40 AA.	
AC	P98185;				
DT	20-AUG-2001 (Rel. 40, Created)				
DT	20-AUG-2001 (Rel. 40, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	RIBOSOME-INACTIVATING PROTEIN SAPORIN-1 (SAP-1) (SO-4) (RRNA N-				
DE	GLYCOSIDASE) (EC 3.2.2.22) (FRAGMENT).				
GN	SAP1.				
OS	Saponaria officinalis (Common soapwort).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Caryophyllales; Caryophyllaceae; Caryophyllaceae; Saponaria.				
OX	NCBI_TaxID=3572;				
RN	[1]				

```

RP SEQUENCE.
RC TISSUE-Leaf;
RX MEDLINE=89336691; PubMed=2753596;
RA Montecuccchi P.-C., Lazzarini A.M., Barbieri L., Stirpe F., Soria M.,
RA Lappi D.;
RT "N-terminal sequence of some ribosome-inactivating proteins.";
RL Int. J. Pept. Protein Res. 33:263-267(1989).
CC -1- FUNCTION: RIBOSOME-INACTIVATING PROTEIN OF TYPE 1, INHIBITS
CC PROTEIN SYNTHESIS IN ANIMAL CELLS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC ADENOSINE ON THE 28S RNA.
CC -1- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
DR InterPro; IPR001574; RIP.
DR PROSITE; PS00275; SHIGA_RICIN; PARTIAL.
KW Protein synthesis inhibitor; Hydrolase; Toxin; Multigene family.
FT NON_TER 40
SQ SEQUENCE 40 AA; 4621 MW; F4615857A9871BF2 CRC64;

Query Match 15.4%; Score 2; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXSD 5
DB 35 GXSD 38

RESULT 3
RBL_CORKO STANDARD; PRT; 466 AA.
ID RBL_CORKO
AC Q3369; 1998 (Rel. 37, Created)
DT 13-DEC-1998 (Rel. 37, Last sequence update)
DT 13-DEC-1998 (Rel. 37, Last annotation update)
DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO
DE LARGE SUBUNIT) (FRAGMENT).
GN RBCL.
OS Cornus kousa.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Cornales; Cornaceae; Cornolideae; Cornus.
OX NCBI_TaxID=28501;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Leaf;
RA Olmstead R.G., Bremer B., Scott K.M., Palmer J.D.;
RT "A parsimony analysis of the Asteridae sensu lato based on rbcl
RT sequences.";
RL Ann. Mo. Bot. Gard. 80:700-722(1993).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
CC 2 3-PHOSPHO-D-GLYCERATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
```

```

CC -----
CC EMBL; LI4395; AAA19755.1;
DR InterPro; IPR000685; RuBisCO_large.
DR Pfam; PF00016; RuBisCO_large; 1.
DR PROSITE; PS00157; RUBISCO_LARGE; PARTIAL.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1
FT ACT_SITE 192
SQ SEQUENCE 466 AA; 51909 MW; 3E27F2E9D798EC69 CRC64;

Query Match 15.4%; Score 2; DB 1; Length 466;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXSD 5
DB 186 GXSD 189

RESULT 4
GRWM_HUMAN STANDARD; PRT; 3 AA.
ID GRWM_HUMAN
AC P01157; 1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE GROWTH-MODULATING PEPTIDE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=77162359; PubMed=858356;
RX Schlesinger D.H., Pickart L., Thaler M.M.;
RA "Growth-modulating serum tripeptide is glycyl-histidyl-lysine.";
RT Experientia 33:324-325(1977).
CC -1- MISCELLANEOUS: THIS SERUM TRIPEPTIDE HAS BEEN FOUND TO STIMULATE
CC GROWTH OF SOME CELL TYPES AND TO INHIBIT OTHER TYPES IN VITRO.
DR PIR; A01421; GHU.
SQ SEQUENCE 3 AA; 340 MW; 6331E81000000000 CRC64;

Query Match 7.7%; Score 1; DB 1; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
DB 1 G 1

RESULT 5
LUXE_VIBFI STANDARD; PRT; 3 AA.
ID LUXE_VIBFI
AC P24272;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE LONG-CHAIN-FATTY-ACID--LUCIFERIN-COMPONENT LIGASE (EC 6.2.1.19)
DE (ACYL-PROTEIN SYNTHETASE) (FRAGMENT).
GN LUXE.
OS Vibrio fischeri.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91072226; PubMed=2254256;
RA Swartzman E., Kapoor S., Graham A.F., Meighen E.A.;
RT "A new Vibrio fischeri lux gene precedes a bidirectional termination
RT site for the lux operon.;
```


RT "Purification and synthesis of eosinophilotoxic tetrapeptides of
 RT human lung tissue: Identification as eosinophil chemotactic factor of
 RL anaphylaxis.";
 CC -1- MISCELLANEOUS: THESE PEPTIDES ARE RELEASED FROM MAST CELLS IN LUNG
 CC (AND OTHER TISSUES) DURING HYPERSENSITIVITY REACTIONS
 CC (ANAPHYLAXIS). THEIR ACTIVITIES, PREFERENTIALLY AFFECTING
 CC EOSINOPHILS, INCLUDE CHEMOTAXIS, CHEMOTACTIC DEACTIVATION, RELEASE
 CC OF ENZYMES, AND STIMULATION OF THE HEXOSE MONOPHOSPHATE SHUNT.
 DR PTR: A03190; ETHUL.
 FT VARIANT 1 1 V -> A (IN OTHER PEPTIDE).
 FT /Frid=VAR_005201.
 SQ SEQUENCE 4 AA; 390 MW; 65058862A0000000 CRC64;

Query Match 7.7%; Score 1; DB 1; Length 4;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
 Db 2 G 2

RESULT 9

ID AL14_CARMA STANDARD; PRT; 5 AA.
 AC P81817; 2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CARCINOSTATIN 14.
 OS Carcinus maenas (Common shore crab) (Green crab).
 CC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 CC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 CC Eubrachyura; Portunoidae; Portunidae; Carcinus.
 OX NCBI_TaxID=6759;
 RN [1]
 RP SEQUENCE.
 RC TISSUE-Cerebral ganglion, and Thoracic ganglion;
 RX MEDLINE=98121193; PubMed=9461295;
 RA Duve H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.P.,
 RA Thorpe A.;
 RT "Isolation and identification of multiple neuropeptides of the
 RT allatostatin superfamily in the shore crab Carcinus maenas.";
 RL Eur. J. Biochem. 250:727-734 (1997).
 CC -1- FUNCTION: MAY ACT AS A NEUROTROSMITTER OR NEUROMODULATOR.
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.
 KW Neuropeptide; Amidation; Multigene family.
 FT MOD_RES 5 5 AMIDATION (POTENTIAL).
 SQ SEQUENCE 5 AA; 586 MW; 672879D5AB300000 CRC64;

Query Match 7.7%; Score 1; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
 Db 4 G 4

RESULT 10

ID BIOA_CITFR STANDARD; PRT; 5 AA.
 AC P13071;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE
 DE (EC 2.6.1.62) (7,8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA
 DE AMINOTRANSFERASE) (FRAGMENT).
 GN BIOA.

OS Citrobacter freundii.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Citrobacter.
 OX NCBI_TaxID=546;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89006280; PubMed=2971595;
 RA Shiuian D., Campbell A.;
 RT "Transcriptional regulation and gene arrangement of Escherichia coli,
 RT Citrobacter freundii and Salmonella typhimurium biotin operons.";
 RL Gene 67:203-211 (1988).
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + 8-AMINO-7-
 CC OXONONANOATE = S-ADENOSYL-4-METHYLTHIO-2-OXOBUTANOATE +
 CC 7,8-DIAMINONONANOATE.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- PATHWAY: BIOTIN BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M21922; ; NOT_ANNOTATED_CDS.
 DR InterPro; IPR000954; Aminotran_3.
 DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; PARTIAL.
 KW Biotin biosynthesis; Transferase; Aminotransferase;
 KW Pyridoxal phosphate.
 FT NON_TER 5 5
 SQ SEQUENCE 5 AA; 582 MW; 6AAAB1B1A6F00000 CRC64;

Query Match 7.7%; Score 1; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 D 5
 Db 4 D 4

RESULT 11

ID BIOA_SALTY STANDARD; PRT; 5 AA.
 AC P12677;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE
 DE (EC 2.6.1.62) (7,8-DIAMINO-PELAGONIC ACID AMINOTRANSFERASE) (DAPA
 DE AMINOTRANSFERASE) (FRAGMENT).
 GN BIOA.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89006280; PubMed=2971595;
 RA Shiuian D., Campbell A.;
 RT "Transcriptional regulation and gene arrangement of Escherichia coli,
 RT Citrobacter freundii and Salmonella typhimurium biotin operons.";
 RL Gene 67:203-211 (1988).
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + 8-AMINO-7-
 CC OXONONANOATE = S-ADENOSYL-4-METHYLTHIO-2-OXOBUTANOATE +
 CC 7,8-DIAMINONONANOATE.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- PATHWAY: BIOTIN BIOSYNTHESIS.

CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M21923; ; NOT_ANNOTATED_CDS.
 DR StyGene; SG10026; bioA.
 DR InterPro; IPR000954; Aminotran_3.
 DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; PARTIAL.
 KW Biotin biosynthesis; Transferase; Aminotransferase;
 KW Pyridoxal phosphate. 5
 FT NON_TER 5
 SQ SEQUENCE 5 AA; 582 MW; 6AAAB1B1A6F00000 CRC64;

Query Match 7.7%; Score 1; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 D 5
 Db 4 D 4

RESULT 12
 ID P2P2_PARMA STANDARD; PRT; 5 AA.
 AC P81864;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PARDAXIN II (PXII) (FRAGMENT).
 OS Pardachirus marmoratus (Red sea moses sole).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;
 CC Soleiidae; Soleidae; Pardachirus.
 OX NCBI_TaxID=31087;
 RN [1]
 RP SEQUENCE.

RC TISSUE-Skin secretion;
 RX MEDLINE=87057369; PubMed=3782138;
 RA Lazarovici P., Primor N., Loew L.M.;
 RT "Purification and pore-forming activity of two hydrophobic
 RT polypeptides from the secretion of the Red sea moses sole (Pardachirus
 RT marmoratus).";
 RL J. Biol. Chem. 261:16704-16713(1986).
 CC -1- FUNCTION: EXHIBITS UNUSUAL SHARK REPELLENT AND SURFACTANT
 CC PROPERTIES. FORMS VOLTAGE-DEPENDENT, ION-PERMEABLE CHANNELS
 CC IN MEMBRANES. AT HIGH CONCENTRATION CAUSES CELL MEMBRANE LYSIS.
 CC -1- SUBUNIT: MONOMER. IN AQUEOUS SOLUTION EXISTS AS A TETRAMER.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE PARDAXIN FAMILY.
 KW Toxin.
 FT NON_TER 5
 SQ SEQUENCE 5 AA; 614 MW; 7769C9C8100000 CRC64;

Query Match 7.7%; Score 1; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 G 2
 Db 1 G 1

RESULT 13
 ID TPIS_CANFA STANDARD; PRT; 5 AA.
 AC P54714;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE TRIOSEPHOSPHATE ISOMERASE (EC 5.3.1.1) (TIM) (FRAGMENT).
 GN TP1.
 OS Canis familiaris (Dog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Heart;
 RX MEDLINE=98163340; PubMed=9504812;
 RA Dunn M.J., Corbett J.M., Wheeler C.H.;
 RT "HSC-2DPAGE and the two-dimensional gel electrophoresis database of
 RT dog heart proteins.";
 RL Electrophoresis 18:2795-2802(1997).
 CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE - DIHYDROXY-
 CC ACETONE PHOSPHATE.
 CC -1- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
 CC HSC-2DPAGE; P54714; DOG.
 DR InterPro; IPR000652; trioseph_isomrse.
 DR PROSITE; PS00171; TIM; PARTIAL.
 DR Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
 KW Pentose shunt.
 FT NON_TER 1 1
 FT NON_TER 5 5
 SQ SEQUENCE 5 AA; 550 MW; 64444862C9A00000 CRC64;

Query Match 7.7%; Score 1; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 G 2
 Db 3 G 3

RESULT 14
 ID TRM3_ECOLI STANDARD; PRT; 5 AA.
 AC P13973;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE TRAM PROTEIN (FRAGMENT).
 GN TRAM.
 OS Escherichia coli.
 CC Plasmid IncFII R100.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88227859; PubMed=2836369;
 RA Inamoto S., Yoshioka Y., Ohtsubo E.;
 RT "Identification and characterization of the products from the traJ
 RT and traY genes of plasmid R100.";
 RL J. Bacteriol. 170:2749-2757(1988).
 CC -1- FUNCTION: TRANSFER GENE PROTEIN. IS INVOLVED IN THE CONJUGATION
 CC PROCESS OF BACTERIAL CELLS FOR THE EXCHANGE OF PLASMID DNA.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: TO TRAM PROTEIN OF OTHER PLASMIDS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M20941; ; NOT_ANNOTATED_CDS.
 DR PIR; A32014; A32014.
 KW Conjugation; Plasmid; DNA-binding.
 FT NON_TER 1
 SQ SEQUENCE 5 AA: 634 MW: 6B1B1AA443500000 CRC64;

Query Match 7.7% Score 1; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 D 5
 Db 3 D 3

RESULT 15
 UF01_MOUSE
 ID UF01_MOUSE STANDARD; PRT: 5 AA.
 AC P38635
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE UNKNOWN PROTEIN FROM 2D-PAGE OF FIBROBLASTS (P19) (FRAGMENT).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Fibroblast;
 RX MEDLINE=95009907; PubMed=7523108;
 RA Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;
 RT "Separation and sequencing of familial and novel murine proteins
 RT using preparative two-dimensional gel electrophoresis.";
 RL Electrophoresis 15:735-745(1994).
 CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
 CC PROTEIN IS: 6.6, ITS MW IS: 19 KDA.
 FT NON_TER 5
 SQ SEQUENCE 5 AA: 717 MW: 7364087043100000 CRC64;

Query Match 7.7% Score 1; DB 1; Length 5;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
 Db 3 G 3

Search completed: January 31, 2002, 13:39:16
 Job time: 76 sec

GénCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:37 ; Search time 130.99 Seconds
(without alignments)
14.517 Million cell updates/sec

Title: US-08-957-709-72
Perfect score: 13
Sequence: 1 XXXXXGXXXXX 13

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0 473505

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_invertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2	15.4	19	11 Q9QV02	Q9QV02 rattus sp.
2	2	15.4	49	11 Q9QV94	Q9QV94 cavia (guin
3	2	15.4	55	5 Q9GW30	Q9GW30 leishmania
4	2	15.4	64	5 Q9GXV2	Q9GXV2 leishmania
5	2	15.4	75	5 Q9GW39	Q9GW39 leishmania
6	2	15.4	92	5 Q9GXV1	Q9GXV1 leishmania
7	2	15.4	99	10 Q64920	Q64920 lycopersico
8	2	15.4	134	5 Q9GX46	Q9GX46 leishmania
9	2	15.4	139	5 Q9GX39	Q9GX39 leishmania
10	2	15.4	148	5 Q9GXR7	Q9GXR7 leishmania
11	2	15.4	166	5 Q9N7N4	Q9N7N4 leishmania
12	2	15.4	192	12 Q90038	Q90038 human immun
13	2	15.4	261	5 Q9GX15	Q9GX15 leishmania
14	2	15.4	290	5 Q9GXY1	Q9GXY1 hylurgops r
15	2	15.4	320	12 Q98Y82	Q98Y82 human immun
16	2	15.4	325	8 Q33491	Q33491 gymnogongru
17	2	15.4	341	12 Q98Z23	Q98Z23 human immun
18	2	15.4	363	8 Q32641	Q32641 mantoniella
19	2	15.4	363	8 Q32870	Q32870 pyramimonas

20	2	15.4	380	8	Q9ZZZ5	Q9ZZZ5 steatomys s
21	2	15.4	382	10	Q9AXK8	Q9AXK8 solanum tub
22	2	15.4	408	8	Q9XNZ4	Q9XNZ4 toxoneuron
23	2	15.4	421	5	Q9NNJ2	Q9NNJ2 leishmania
24	2	15.4	466	8	Q96082	Q96082 stephanosia
25	2	15.4	487	2	P94205	P94205 agrobacteri
26	2	15.4	603	5	Q9GWR1	Q9GWR1 leishmania
27	1	7.7	5	13	P82070	P82070 litoria rub
28	1	7.7	5	13	P82073	P82073 litoria rub
29	1	7.7	7	2	O07354	O07354 synechococc
30	1	7.7	7	2	O47029	O47029 enterobacte
31	1	7.7	7	2	O47505	O47505 escherichia
32	1	7.7	7	6	O28742	O28742 oryctolagus
33	1	7.7	7	10	O49223	O49223 glycine max
34	1	7.7	7	10	P82445	P82445 nicotiana t
35	1	7.7	7	10	Q9C5B3	Q9C5B3 arabidopsis
36	1	7.7	7	11	O63668	O63668 rattus norv
37	1	7.7	7	11	O63480	O63480 rattus norv
38	1	7.7	7	11	O55184	O55184 rattus norv
39	1	7.7	7	12	O07624	O07624 rous sarcom
40	1	7.7	8	2	O53790	O53790 streptococc
41	1	7.7	8	2	O09258	O09258 synechococc
42	1	7.7	8	2	O56140	O56140 streptococc
43	1	7.7	8	2	O56429	O56429 thermus aqu
44	1	7.7	8	2	P72279	P72279 rhodococcus
45	1	7.7	8	2	O56759	O56759 xanthobacte

ALIGNMENTS

RESULT 1
ID Q9QV02 PRELIMINARY; PRT; 19 AA.
AC Q9QV02;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE C-REACTIVE PROTEIN (FRAGMENT).
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118;
RN [1]
RP SEQUENCE.
RX MEDLINE=95355312; PubMed=7629028;
RA Motojima K., Goto S.;
RL J. Biochem. 117:597-602(1995).
SQ SEQUENCE 19 AA; 2184 MW; A53237547BE09F0D CRC64;

Query Match 15.4%; Score 2; DB 11; Length 19;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GXXD 5
Db 13 GXXD 16
RESULT 2
ID Q9QV94 PRELIMINARY; PRT; 49 AA.
AC Q9QV94;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE ORGAN OF CORTI PROTEIN (FRAGMENT).
OS Cavia (guinea pigs).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae.
OX NCBI_TaxID=10140;
RN [1]

RP SEQUENCE.
 RA MEDLINE-93163013; PubMed-8432690;
 RX Thalmann I., Suzuki H., McCourt D.W., Comegys T.H., Thalmann R.;
 FT "Partial amino acid sequences of organ of Corti proteins OCP1 and
 RT OCP2: progress report".
 RL Hear. Res. 64:191-198(1993).
 SQ SEQUENCE 49 AA; 5327 MW; A90E7FE21BB986CA CRC64;

Query Match 15.4%; Score 2; DB 11; Length 49;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
 Db 10 GXXD 13

RESULT 3
 Q9GW30 PRELIMINARY; PRT; 55 AA.
 AC Q9GW30
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE HYPOTHETICAL 6.1 KDA PROTEIN (FRAGMENT).
 GN LM12.1101
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
 RA Oliver K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL390114; CAC02703.1;
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 55
 SQ SEQUENCE 55 AA; 6092 MW; A325DA28305F9331 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 55;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
 Db 32 GXXD 35

RESULT 4
 Q9GXV2 PRELIMINARY; PRT; 64 AA.
 AC Q9GXV2
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE HYPOTHETICAL 7.0 KDA PROTEIN (FRAGMENT).
 GN LM12.231.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
 RA Oliver K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL390114; CAC02096.1;
 DR InterPro: IPR001208; MCM;
 DR PROSITE: PS50051; MCM_2; 1.

KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 64
 SQ SEQUENCE 64 AA; 7016 MW; 515441B7AFD6EBBA CRC64;

Query Match 15.4%; Score 2; DB 5; Length 64;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
 Db 36 GXXD 39

RESULT 5
 Q9GW39 PRELIMINARY; PRT; 75 AA.
 AC Q9GW39
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE HYPOTHETICAL 7.8 KDA PROTEIN (FRAGMENT).
 GN LM12.1088
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
 RA Oliver K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL390114; CAC02694.1;
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 75
 SQ SEQUENCE 75 AA; 7754 MW; 507158F8252B1A42 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
 Db 67 GXXD 70

RESULT 6
 Q9GXV1 PRELIMINARY; PRT; 92 AA.
 AC Q9GXV1
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE HYPOTHETICAL 10.0 KDA PROTEIN (FRAGMENT).
 GN LM12.232.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
 RA Oliver K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL390114; CAC02097.1;
 DR InterPro: IPR001208; MCM;
 DR PROSITE: PS50051; MCM_2; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 92

```
SQ SEQUENCE 92 AA; 9991 MW; BE1E32EA37C402B9 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 92;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
DB 32 GXND 35

RESULT 7
O64920 PRELIMINARY; PRT; 99 AA.
AC O64920;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE IAA9 (FRAGMENT).
GN IAA9.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ETIOLATED HYPOCOTYLS;
RA Nebenfuhr A., Lomax T.L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020202; AAC13260.1; -
DR Mendel; 28774; Lyces:1524; 28774.
DR InterPro; IPR003311; AUX_IAA.
DR Pfam; PF02309; AUX_IAA; 1.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 10767 MW; 9F9F01A3EBCAF243 CRC64;

Query Match 15.4%; Score 2; DB 10; Length 99;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
DB 75 GXND 78

RESULT 8
O9GX46 PRELIMINARY; PRT; 134 AA.
AC O9GX46;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 14.3 KDA PROTEIN (FRAGMENT).
GN LM12_571.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAC02345.1; -
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 134
SQ SEQUENCE 134 AA; 14304 MW; F39B0EC2A8D22A4 CRC64;

SQ SEQUENCE 92 AA; 9991 MW; BE1E32EA37C402B9 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 92;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
DB 32 GXND 35

RESULT 7
O64920 PRELIMINARY; PRT; 99 AA.
AC O64920;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE IAA9 (FRAGMENT).
GN IAA9.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ETIOLATED HYPOCOTYLS;
RA Nebenfuhr A., Lomax T.L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020202; AAC13260.1; -
DR Mendel; 28774; Lyces:1524; 28774.
DR InterPro; IPR003311; AUX_IAA.
DR Pfam; PF02309; AUX_IAA; 1.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 10767 MW; 9F9F01A3EBCAF243 CRC64;

Query Match 15.4%; Score 2; DB 10; Length 99;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
DB 75 GXND 78

RESULT 8
O9GX46 PRELIMINARY; PRT; 134 AA.
AC O9GX46;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 14.3 KDA PROTEIN (FRAGMENT).
GN LM12_571.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAC02345.1; -
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 134
SQ SEQUENCE 134 AA; 14304 MW; F39B0EC2A8D22A4 CRC64;

SQ SEQUENCE 92 AA; 9991 MW; BE1E32EA37C402B9 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 134;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
DB 58 GXND 61

RESULT 9
O9GX39 PRELIMINARY; PRT; 139 AA.
AC O9GX39;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 15.1 KDA PROTEIN (FRAGMENT).
GN LM12_584.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAC02352.1; -
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15117 MW; 084D966737B6BFBB CRC64;

Query Match 15.4%; Score 2; DB 5; Length 139;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
DB 109 GXND 112

RESULT 10
O9GX7 PRELIMINARY; PRT; 148 AA.
AC O9GX7;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 15.2 KDA PROTEIN (FRAGMENT).
GN LM12_272.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAC02129.1; -
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 148
SQ SEQUENCE 148 AA; 15210 MW; DC9CDEEECEEE9EC1 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 148;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 GXXD 5
Db 74 GXXD 77

RESULT 11
Q9N7N4 ID Q9N7N4 PRELIMINARY; PRT; 166 AA.
AC Q9N7N4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POSSIBLE NUCLEOLAR TRANSCRIPTION FACTOR 2 (FRAGMENT).
GN L28.170.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL390935; CAC00903.1; -.
FT NON_TER 1
FT NON_TER 166
SQ SEQUENCE 166 AA; 17573 MW; 111FEE2C415FB6C4 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 166;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
Db 154 GXXD 157

RESULT 12
O90038 ID O90038 PRELIMINARY; PRT; 192 AA.
AC O90038;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M2;
RX MEDLINE=58248614; PubMed=9582120;
RA Frenkel L.M., Mullins J.I., Learn G.H., Manns-Arcuino L.,
Harrington B.L., Kallish M.L., Steketee R.W., Thea D.M., Nichols J.E.,
Liu S.-L., Harmache A., He X., Muthui D., Madan A., Hood L.,
Haase A.T., Zupan M., Staskus K., Wolinsky S.M., Krostad P.,
Zhao J.-O., Chen I., Koup R., Ho D.D., Korber B.T., Apple R.J.,
Coombs R.W., Pahwa S., Roberts N.J. Jr.;
RT "Genetic evaluation of suspected cases of transient HIV-1 infection of
infants."
RL Science 280:1073-1077(1998).
DR EMBL: AF065590; AAC21548.1; -.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
KW Envelope protein.
FT NON_TER 1
FT NON_TER 192
SQ SEQUENCE 192 AA; 21540 MW; 95E23174393B1028 CRC64;

Query Match 15.4%; Score 2; DB 12; Length 192;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
Db 133 GXXD 136

RESULT 13
Q9GX15 ID Q9GX15 PRELIMINARY; PRT; 261 AA.
AC Q9GX15;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HYPOTHETICAL 27.9 KDA PROTEIN (FRAGMENT).
GN LM12.383.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
Oliviera K., (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted;
DR EMBL: AL390114; CAC02208.1; -.
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 261
SQ SEQUENCE 261 AA; 27941 MW; E1C505B36BEA9832 CRC64;

Query Match 15.4%; Score 2; DB 5; Length 261;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DXXXG 9
Db 171 DXXXG 175

RESULT 14
O9GPLY ID O9GPLY PRELIMINARY; PRT; 290 AA.
AC O9GPLY;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ELONGATION FACTOR 1 ALPHA (FRAGMENT).
GN EF-1A.
OS Hylurgops rugipennis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cuculiformia; Phytophaga; Scolytidae; Hylurgops.
OX NCBI_TaxID=141166;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HLH04;
RA Sequeira A.S., Normark B.B., Farrell B.D.;
RT "Evolutionary assembly of the conifer fauna: Distinguishing ancient
from recent associations in bark beetles."
RL Proc. R. Soc. Lond., B, Biol. Sci. 0:0-0(2001).
CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
DR EMBL: AF308408; AAC45067.1; -.
DR InterPro; IPR000795; GTP_EFTU.
DR Pfam; PF00009; GTP_EFTU; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR PROSITE; PS00301; EFACOR_GTP; 1.
KW GTP-binding; Protein biosynthesis.
FT NON_TER 1
FT NON_TER 290
SQ SEQUENCE 290 AA; 31705 MW; 32CD038BA321A628 CRC64;
```

Query Match 15.4%; Score 2; DB 5; Length 290;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
 ||||
DB 40 GXXD 43

RESULT 15

Q98Y82 PRELIMINARY; PRT; 320 AA.
AC Q98Y82;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POL PROTEIN (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=997065;
RA Schmidt B., Walter H., Moschik G., Paatz C., Werwein M., Schwingel E.,
RA Korn K.;
RT "Recovery of HIV-1 pol gene sequences by direct sequencing of
RT amplification products derived from plasma samples";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF347587; AAK32663.1; -.
FT NON_TER 1
FT NON_TER 320
SQ SEQUENCE 320 AA; 36480 MW; 89F1BA75FE57B37A CRC64;

Query Match 15.4%; Score 2; DB 12; Length 320;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
 ||||
DB 27 GXXD 30

Search completed: January 31, 2002, 13:37:38
Job time: 163 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:08 ; Search time 78.64 Seconds
(without alignments)
12.592 Million cell updates

```

Title:      US-08-957-709-72
Perfect score: 13
Sequence:   1 XGXXDXXXGXXXX 13

Scoring table:  OLIGO

```

```
Searched:      219241 seqs, 76174552 residues
Word size :    0
Total number of hits satisfying chosen parameters: 219241
```

```

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Listing first 45 summaries

```

```
Database :      PIR_68:*
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2	15.4	28	2	A60698	trichocyst protein	
2	2	15.4	34	2	S65417	pyruvate synthase	
3	2	15.4	40	2	S6331	rRNA N-glycosidase	
4	2	15.4	99	2	T05708	auxin-induced prot	
5	2	15.4	158	1	GCPG	glucagon precursor	
6	2	15.4	254	2	B31790	Ig heavy chain V r	
7	1	7.7	3	2	GKHU	growth-modulating	
8	1	7.7	3	2	A60898	bursin - chicken	
9	1	7.7	3	2	A23751	spinal cord peptid	
10	1	7.7	3	2	B23751	spinal cord peptid	
11	1	7.7	3	2	PT0636	T-cell receptor be	
12	1	7.7	3	2	PT0571	T-cell receptor be	
13	1	7.7	4	1	ECXAA	antho-RFamide neur	
14	1	7.7	4	2	A32039	tyrosine-melanocyt	
15	1	7.7	4	2	PL0140	carbon-monoxide de	
16	1	7.7	4	2	I40697	biotin A - Citroba	
17	1	7.7	4	2	A41890	protein D - Escher	
18	1	7.7	4	2	D41654	hypothetical prote	
19	1	7.7	4	2	S09478	globulin IV alpha	
20	1	7.7	4	2	T30569	hypothetical prote	
21	1	7.7	4	2	I38888	COI intron 16 prot	
22	1	7.7	4	2	A25844	autho-RF amide neu	
23	1	7.7	4	2	A34626	RPCH-related neuro	
24	1	7.7	4	2	A32480	achatin-I - giant	
25	1	7.7	4	2	I51049	metallothionein-A	
26	1	7.7	4	2	S93390	myosin-light-chain	
27	1	7.7	4	2	PT0240	Ig heavy chain CRD	
28	1	7.7	4	2	PT0271	Ig heavy chain CRD	
29	1	7.7	4	2	S43959	Ig mu chain V regi	

ALIGNMENTS

RESULT 1

A60698
trichocyst protein 12 - Paramesidium tetraurelia (fragment)
C:Species: Paramesidium tetraurelia
C:Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 07-Dec-1999
C:Accession: A60698
R;Findall, S.H.; Devito, L.D.; Nelson, D.L.
J. Cell Sci. 92, 441-447, 1989
A:Title: Biochemical characterization of the proteins of Paramesidium secretory granule
A:Reference number: A60698; MUID:90078398
A:Accession: A60698
A:Molecule type: protein
A:Residues: 1-28 <TIN>
C:Comment: This protein was purified from the extruded matrix (contents) of trichocysts
C:Genetics:
A:Genetic code: SGC5
C:Keywords: extracellular protein

Query Match	15.4%	Score 2;	DB 2;	Length 28;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4:	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

Qy	2	GXXD	5
Dp	22	GXXD	25

DEC 11 1967

pyruvate synthase (EC 1.2.7.1) beta chain - Methanosarcina barkeri (fragment)
 N:Alternate names: pyruvate:ferredoxin 2-oxido-reductase (CoA-acetylating) beta chain
 C:Species: Methanosarcina barkeri
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 05-May-2000
 C:Accession: S65417
 R:Bock, A.K.; Kunow, J.; Glasenmacher, J.; Schoenheit, P.
 Eur. J. Biochem. 237, 35-44, 1996
 A:Title: Catalytic properties, molecular composition and sequence alignments of pyruvate
 A:Reference number: S65416; MUID:96203906
 A:Accession: S65417
 A:Molecule type: protein
 A:Residues: 1-34 <BOC>
 A:Experimental source: strain Fusaro
 C:Keywords: coenzyme A; oxidoreductase

Query Match	15.4%	Score 2;	DB 2;	Length 34;
Best Local Similarity	100.0%;	pred. NO. 0;		
Matches	4;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	

Qy 2 GXXD 5
||||

Db 18 GXXD 21

RESULT 3

S16331
rRNA N-glycosidase (EC 3.2.2.22) 1 - common soapwort (fragment)
N:Alternate names: ribosome-inactivating protein SO-4; saporin 1
C:Species: Saponaria officinalis (common soapwort)
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 23-Jun-1993
C:Accession: S16331
R:Montecucchi, P.C.; Lazarini, A.M.; Barbieri, L.; Stirpe, F.; Soria, M.; Lappi, D.
Int. J. Pept. Protein Res. 33: 263-267, 1989
A:Title: N-terminal sequence of some ribosome-inactivating proteins.
A:Reference number: S16331; MUID:89326691
A:Accession: S16331
A:Molecule type: Protein
A:Residues: 1-40 <NON>
C:Keywords: glycosidase; hydrolase

Query Match 15.4%; Score 2; DB 2; Length 40;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 0;

QY 2 GXXD 5

Db 35 GXXD 38

RESULT 4

T05708
auxin-induced protein IAA9 - tomato (fragment)
N:Species: Lycopersicon esculentum (tomato)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Aug-1999
C:Accession: T05708
R:Nebenfuhr, A.; Lomax, T. L.
submitted to the EMBL Data Library, September 1997
A:Description: Auxin-regulated LeIAA genes of tomato.
A:Reference number: Z15303
A:Accession: T05708
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-99 <NEP>
A:CROSS-references: EMBL:AF022020; NID:g3043908; PIDN:AA013260.1; PID:g3043909
A:Experimental source: tissue type etiolated hypocotyls
C:Genetics:
A:Gene: IAA9
C:Superfamily: auxin-induced protein aux28

Query Match 15.4%; Score 2; DB 2; Length 99;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 0;

QY 2 GXXD 5

Db 75 GXXD 78

RESULT 5

GCPC
glucagon precursor - pig (fragment)
N:Alternate names: glucicentin; oxyntomodulin
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 17-Dec-1982 #sequence_revision 31-Mar-1993 #text_change 20-Mar-1998
C:Accession: A01540; A60312; A91781; B32614; A28064
R:Thim, L.; Moody, A. J.
Regul. Pept. 2: 139-150, 1981
A:Title: The primary structure of porcine glucicentin (proglucagon).
A:Reference number: A94233; MUID:81248172
A:Accession: A01540
A:Molecule type: protein

A:Residues: 1-69 <THL>
R:Thim, L.; Moody, A. J.
Regul. Pept. Suppl. 2, S33, 1983
A:Title: Primary structure of a possible porcine proglucagon fragment.
A:Reference number: A60312
A:Accession: A60312
A:Molecule type: protein
A:Residues: 1-30 <TH2>
A:Note: this peptide is co-secreted with glucagon from the pancreas
R:Bromer, W. W.; Sinn, L. G.; Behrens, O. K.
J. Am. Chem. Soc. 79, 2807-2810, 1957
A:Title: The amino acid sequence of glucagon. V. Location of amide groups, acid degra
A:Reference number: A91781
A:Accession: A91781
A:Molecule type: protein
A:Residues: 35-61 <BRO>
R:Orskov, C.; Bersani, M.; Johnsen, A. H.; Hojrup, P.; Holst, J. J.
J. Biol. Chem. 264, 12826-12829, 1989
A:Title: Complete sequences of glucagon-like peptide-1 from human and pig small intes
A:Reference number: A92752; MUID:89327238
A:Accession: B32614
A:Molecule type: protein
A:Residues: 78-107 <ORS>
R:Buhl, T.; Thim, L.; Kofod, H.; Orskov, C.; Harling, H.; Holst, J. J.
J. Biol. Chem. 263, 8621-8624, 1988
A:Title: Naturally occurring products of proglucagon 111-160 in the porcine and human
A:Reference number: A28064; MUID:88243712
A:Accession: A28064
A:Molecule type: protein
A:Residues: 111-158 <BUH>
C:Comment: X's represent missing amino acids, mostly basic, that are predicted to exi
C:Superfamily: glucagon
C:Keywords: amidated carboxyl end; carbohydrate metabolism; duplication; hormone; int
F:1-69/Product: glucagon-69 #status experimental <G69>
F:1-30/Region: glucicentin-related peptide #status experimental
F:33-69/Product: glucagon-37 #status predicted <G37>
F:33-61/Product: glucagon #status experimental <G37>
F:78-107/Product: glucagon-like peptide 1 #status experimental <GL1>
F:126-158/Product: glucagon-like peptide 2 #status experimental <GL2>
F:107/Modified site: amidated carboxyl end (Arg) (amide in mature form from following

Query Match 15.4%; Score 2; DB 1; Length 158;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 0;

QY 2 GXXD 5

Db 108 GXXD 111

RESULT 6

B31790
Ig heavy chain V region (17/9) - mouse
C:Species: Mus musculus (house mouse)
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 23-May-1997
C:Accession: B31790
R:Schulze-Gahmen, U.; Rini, J. M.; Arevalo, J.; Stura, E. A.; Kenten, J. H.; Wilson, I. A
J. Biol. Chem. 263, 17100-17105, 1988
A:Title: Preliminary crystallographic data, primary sequence, and binding data for an
A:Reference number: A92686; MUID:89034213
A:Accession: B31790
A:Molecule type: mRNA
A:Residues: 1-254 <SCH>
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: heterotrimer; immunoglobulin
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 15.4%; Score 2; DB 2; Length 254;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 4; Conservative 0;

QY 2 GXXD 5
|||||
Db 136 GXXD 139

RESULT 7
GKHU
growth-modulating peptide - human
C:Species: Homo sapiens (man)
C:Date: 23-Oct-1981 #sequence_revision 26-May-1995 #text_change 20-Jun-2000
C:Accession: A01421
R:Schlesinger, D.H.; Pickart, L.; Thaler, M.M.
Experientia 33, 324-325, 1977
A:Title: Growth-modulating serum tripeptide is glycyl-histidyl-lysine.
A:Reference number: A01421; MUID:77162369
A:Accession: A01421
A:Molecule type: protein
A:Residues: 1-3 <SCH>
A:Note: this serum tripeptide is found to stimulate growth of some cell types and to inhibit bursin production in chicken
C:Superfamily: unassigned animal peptides

Query Match 7.7%; Score 1; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
|
Db 1 G 1

RESULT 8
A60898
bursin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 18-Aug-2000
C:Accession: A60898
R:Audhya, T.; Kroon, D.; Heavner, G.; Viamontes, G.; Goldstein, G.
Science 231, 997-999, 1986
A:Title: Tripeptide structure of bursin, a selective B-cell-differentiating hormone of the bursin-producing B-cell line B-220/1
A:Reference number: A60898; MUID:86122916
A:Accession: A60898
A:Molecule type: protein
A:Residues: 1-3 <AUD>
C:Superfamily: unassigned animal peptides
C:Keywords: amidated carboxyl end; hormone
F:3/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 7.7%; Score 1; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
|
Db 3 G 3

RESULT 9
A23751
spinal cord peptide SCP-4 - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 28-Sep-1987 #sequence_revision 28-Sep-1987 #text_change 18-Aug-2000
C:Accession: A23751
R:Hsi, K.L.; Chen, R.L.; Chen, Z.G.; Zhang, H.L.; Lu, Y.A.; Guo, S.Y.; Wu, S.X.; Tsou, K.
Arch. Biochem. Biophys. 240, 178-183, 1985
A:Reference number: A23751; MUID:85250425
A:Accession: A23751
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-3 <HSI>
C:Superfamily: unassigned animal peptides

Query Match 7.7%; Score 1; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
|
Db 2 G 2

RESULT 10
B23751
spinal cord peptide SCP-5 - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 28-Sep-1987 #sequence_revision 28-Sep-1987 #text_change 18-Aug-2000
C:Accession: B23751
R:Hsi, K.L.; Chen, R.L.; Chen, Z.G.; Zhang, H.L.; Lu, Y.A.; Guo, S.Y.; Wu, S.X.; Tsou, K.
Arch. Biochem. Biophys. 240, 178-183, 1985
A:Reference number: A23751; MUID:85250425
A:Accession: B23751
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-3 <HSI>
C:Superfamily: unassigned animal peptides

Query Match 7.7%; Score 1; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
|
Db 3 G 3

RESULT 11
PT0636
T-cell receptor beta chain V-D-J region (100-2AT) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C:Accession: PT0636
R:Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A:Reference number: PT0509; MUID:91277601
A:Accession: PT0636
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3 <FEE>
A:Experimental source: newborn thymus, strain BALB/c
C:Keywords: T-cell receptor

Query Match 7.7%; Score 1; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 G 2
|
Db 3 G 3

RESULT 12
PT0571
T-cell receptor beta chain V-D-J region (141-1CM) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 30-May-1997
C:Accession: PT0571
R:Feeney, A.J.
J. Exp. Med. 174, 115-124, 1991
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions
A:Reference number: PT0509; MUID:91277601
A:Accession: PT0571

A:Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-3 <PEE>

C:Experimental source: day 19 fetal thymus, strain BALB/c

C:Keywords: T-cell receptor

Query Match 7.7%: Score 1; DB 2; Length 3;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 G 2

Db 3 G 3

RESULT 13

ECXAA

C:Species: Anthopleura elegantissima

C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 08-Dec-1995

C:Accession: A26666

R:Grimmelikhuizen, C. J. P.; Graff, D.

Proc. Natl. Acad. Sci. U.S.A. 83, 9817-9821, 1986

A:Title: Isolation of <Glu-Gly-Arg-Phe-NR2 (Antho-RFamide), a neuropeptide from sea anem

A:Reference number: A26666; MUID:87092339

A:Accession: A26666

A:Molecule type: protein

A:Residues: 1-4 <GRI>

C:Comment: The function of this peptide is not known but it could act as a transmitter a

C:Comment: Synthetic and natural peptides had identical properties.

C:Superfamily: RFamide neuropeptide

C:Keywords: amidated carboxyl end; neuropeptide; pyroglutamic acid

F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

F:4/Modified site: amidated carboxyl end (Phe) #status experimental

Query Match 7.7%: Score 1; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 G 2

Db 2 G 2

RESULT 14

A32039

C:Species: Bos primigenius taurus (cattle)

C:Date: 31-Jul-1989 #sequence_revision 31-Jul-1989 #text_change 18-Aug-2000

C:Accession: A32039

R:Horvath, A.; Kastin, A. J.

J. Biol. Chem. 264, 2175-2179, 1989

A:Title: Isolation of tyrosine-melanocyte-stimulating hormone release-inhibiting factor

A:Reference number: A32039; MUID:89123285

A:Accession: A32039

A:Molecule type: protein

A:Residues: 1-4 <HOR>

A:Experimental source: brain

C:Superfamily: unassigned animal peptides

C:Keywords: amidated carboxyl end

F:4/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 7.7%: Score 1; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 G 2

Db 4 G 4

RESULT 15

PL0140

C:Species: Pseudomonas carboxydohydrog

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 28-Apr-1993

C:Accession: PL0140

R:Kraut, M.; Hugendieck, I.; Herwig, S.; Meyer, O.

Arch. Microbiol. 152, 335-341, 1989

A:Title: Homology and distribution of CO dehydrogenase structural genes in carboxydot

A:Reference number: PL0138; MUID:90055678

A:Accession: PL0140

A:Molecule type: protein

A:Residues: 1-4 <KRA>

C:Comment: Carbon-monoxide dehydrogenase consists of three polypeptide chains: large,

C:Keywords: oxidoreductase

Query Match 7.7%: Score 1; DB 2; Length 4;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 G 2

Db 2 G 2

Search completed: January 31, 2002, 13:20:09

Job time: 110 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:10 ; Search time 65.13 seconds
(without alignments)
4.837 Million cell updates/sec

Title: US-08-957-709-73
Perfect score: 14
Sequence: 1 FAWVDPGWDGNTLM 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	4	US-08-822-774-54
2	11	78.6	246	4	Sequence 54, Appl
3	6	42.9	561	2	Sequence 44, Appl
4	5	35.7	13	2	Sequence 12, Appl
5	5	35.7	13	2	Sequence 98, Appl
6	5	35.7	13	2	Sequence 98, Appl
7	5	35.7	13	2	Sequence 98, Appl
8	5	35.7	13	2	Sequence 98, Appl
9	5	35.7	13	3	Sequence 98, Appl
10	5	35.7	13	3	Sequence 98, Appl
11	5	35.7	13	4	Sequence 98, Appl
12	5	35.7	13	4	Sequence 98, Appl
13	5	35.7	16	2	Sequence 7, Appl
14	5	35.7	16	4	Sequence 7, Appl
15	5	35.7	16	4	Sequence 7, Appl
16	5	35.7	27	1	Sequence 20, Appl
17	5	35.7	27	2	Sequence 20, Appl
18	5	35.7	27	2	Sequence 20, Appl
19	5	35.7	27	4	Sequence 20, Appl
20	5	35.7	143	4	Sequence 10, Appl
21	5	35.7	171	1	Sequence 2, Appl
22	5	35.7	172	4	Sequence 7, Appl
23	5	35.7	175	4	Sequence 3, Appl
24	5	35.7	200	1	Sequence 1, Appl
25	5	35.7	200	1	Sequence 1, Appl
26	5	35.7	207	1	Sequence 3, Appl
27	5	35.7	207	1	Sequence 3, Appl

28	5	35.7	210	2	US-08-726-136-1	Sequence 1, Appl
29	5	35.7	210	4	US-09-103-434-1	Sequence 1, Appl
30	5	35.7	210	4	US-09-687-594-1	Sequence 1, Appl
31	5	35.7	234	4	US-09-354-129-14	Sequence 14, Appl
32	5	35.7	250	2	US-08-872-961A-1	Sequence 1, Appl
33	5	35.7	250	3	US-09-231-258-1	Sequence 1, Appl
34	5	35.7	326	6	5171684-7	Patent No. 5171684
35	5	35.7	337	1	US-08-440-856A-3	Sequence 3, Appl
36	5	35.7	418	2	US-08-305-764C-58	Sequence 58, Appl
37	5	35.7	418	2	US-08-305-764C-60	Sequence 60, Appl
38	5	35.7	557	1	US-08-313-288B-16	Sequence 16, Appl
39	5	35.7	560	2	US-08-559-492-5	Sequence 5, Appl
40	5	35.7	613	1	US-08-272-875-3	Sequence 3, Appl
41	5	35.7	647	2	US-08-305-764C-56	Sequence 56, Appl
42	5	35.7	1968	1	US-07-745-206A-7	Sequence 7, Appl
43	5	35.7	1968	1	US-08-455-543A-45	Sequence 45, Appl
44	5	35.7	1968	2	US-08-223-305C-45	Sequence 45, Appl
45	5	35.7	1968	2	US-08-311-363-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-822-774-54
; Sequence 54, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-08-822-774-54

Query Match 100.0%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 9.1e-10;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FAWVDPGWDGNTLM 14

```
Db      1 FAWVDPGWDGNTLM 14
RESULT 2
US-08-822-774-44
; Sequence 44, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; EXTREASES: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8840
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
US-08-822-774-44

Query Match 78.68; Score 11; DB 4; Length 246;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 FAWVDPGWDGN 11
Db      39 FAWVDPGWDGN 49
RESULT 3
US-08-559-492-12
; Sequence 12, Application US/08559492
; Patent No. 5843884
; GENERAL INFORMATION:
; APPLICANT: Sims, Peter J.
; TITLE OF INVENTION: C9 Complement Inhibitor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center, 1201 West Peachtree
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,492
; FILING DATE: 15-NOV-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-873-8794
; TELEFAX: 404-873-8795
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-559-492-12

Query Match 42.9%; Score 6; DB 2; Length 561;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 WDGNTL 13
Db      190 WDGNTL 195
RESULT 4
US-08-934-222-98
; Sequence 98, Application US/08934222
; Patent No. 5928896
; GENERAL INFORMATION:
; APPLICANT: EVANS, Herbert J.
; APPLICANT: KINI, R. Manjunatha
; TITLE OF INVENTION: Polypeptides That Include Conformation-
; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti
; TITLE OF INVENTION: Site
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,222
; FILING DATE: 19-SEPT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/532,818
; FILING DATE: 03-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/143,364
; FILING DATE: 29-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/051,741
; FILING DATE: 23-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Isacson, John P.
; REGISTRATION NUMBER: 33,751
; REFERENCE/DOCKET NUMBER: 040433/0148
```

INFORMATION FOR SEQ ID NO: 98:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 13 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-934-222-98

Query Match 35.7%; Score 5; DB 2; Length 13;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
 DB 7 DGNTL 11

RESULT 5
 US-08-933-402-98
 ; Sequence 98, Application US/08933402
 ; Patent No. 5948887
 ; GENERAL INFORMATION:
 ; APPLICANT: EVANS, Herbert J.
 ; APPLICANT: KINI, R. Manjunatha
 ; TITLE OF INVENTION: Polypeptides That Include Conformation-
 ; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: Suite 500, 3000 K Street NW
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20007

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/933,402
 FILING DATE: 19-SEPT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/532,818
 FILING DATE: 03-MAY-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: U.S. 08/051,741
 FILING DATE: 29-OCT-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Isacson, John P.
 REGISTRATION NUMBER: 33,751
 REFERENCE/DOCKET NUMBER: 040433/0148
 INFORMATION FOR SEQ ID NO: 98:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 13 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-933-402-98

Query Match 35.7%; Score 5; DB 2; Length 13;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
 DB 7 DGNTL 11

RESULT 6
 US-09-207-621-98
 ; Sequence 98, Application US/09207621
 ; Patent No. 5952465
 ; GENERAL INFORMATION:
 ; APPLICANT: EVANS, Herbert J.
 ; APPLICANT: KINI, R. Manjunatha
 ; TITLE OF INVENTION: Polypeptides That Include Conformation-
 ; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interactio
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: Suite 500, 3000 K Street NW
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20007

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/207,621
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/532,818
 FILING DATE: 03-MAY-1996
 APPLICATION NUMBER: PCT/US94/04294
 FILING DATE: 21-APR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: U.S. 08/143,364
 FILING DATE: 29-OCT-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: U.S. 08/051,741
 FILING DATE: 23-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Isacson, John P.
 REGISTRATION NUMBER: 33,751
 REFERENCE/DOCKET NUMBER: 040433/0148
 INFORMATION FOR SEQ ID NO: 98:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 13 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-09-207-621-98

Query Match 35.7%; Score 5; DB 2; Length 13;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
 DB 7 DGNTL 11

RESULT 7
 US-08-532-818-98
 ; Sequence 98, Application US/08532818
 ; Patent No. 5965698
 ; GENERAL INFORMATION:
 ; APPLICANT: EVANS, Herbert J.
 ; APPLICANT: KINI, R. Manjunatha
 ; TITLE OF INVENTION: Polypeptides That Include Conformation-
 ; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: Suite 500, 3000 K Street NW
 ; CITY: Washington
 ; STATE: DC

FILING DATE: 29-OCT-1993

Query Match 35.7%; Score 5; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
Db 7 DGNTL 11

RESULT 10
US-08-933-843-98
; Sequence 98, Application US/08933843
; Patent No. 611069
; GENERAL INFORMATION:
; APPLICANT: EVANS, Herbert J.
; APPLICANT: KINI, R. Manjunatha
; TITLE OF INVENTION: Polypeptides That Include Conformation-
; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interaction
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,843
; FILING DATE: 19-SEPT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/532,818
; FILING DATE: 03-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/143,364
; FILING DATE: 29-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/051,741
; FILING DATE: 23-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Isaacson, John P.
; REGISTRATION NUMBER: 33,751
; REFERENCE/DOCKET NUMBER: 040433/0148
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-933-843-98

Query Match 35.7%; Score 5; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
Db 7 DGNTL 11

RESULT 11
US-08-934-223-98
; Sequence 98, Application US/08934223
; Patent No. 6147189
; GENERAL INFORMATION:
; APPLICANT: EVANS, Herbert J.
; APPLICANT: KINI, R. Manjunatha
; TITLE OF INVENTION: Polypeptides That Include Conformation-

; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti
; TITLE OF INVENTION: Site
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,223
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,818
; FILING DATE: 03-MAY-1996
; APPLICATION NUMBER: PCT/US94/04294
; FILING DATE: 21-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/143,364
; FILING DATE: 29-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/051,741
; FILING DATE: 23-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Isaacson, John P.
; REGISTRATION NUMBER: 33,751
; REFERENCE/DOCKET NUMBER: 040433/0148
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-934-223-98

Query Match 35.7%; Score 5; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
Db 7 DGNTL 11

RESULT 12
US-09-413-492-98
; Sequence 98, Application US/09413492
; Patent No. 6258550
; GENERAL INFORMATION:
; APPLICANT: EVANS, Herbert J.
; APPLICANT: KINI, R. Manjunatha
; TITLE OF INVENTION: Polypeptides That Include Conformation-
; TITLE OF INVENTION: Constraining Groups Which Flank A Protein-Protein Interacti
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/413,492
; FILING DATE:
; PRIOR APPLICATION NUMBER: 08/532,818
; FILING DATE: 03-MAY-1996
; APPLICATION NUMBER: PCT/US94/04294
; FILING DATE: 21-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/143,364
; FILING DATE: 29-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/051,741
; FILING DATE: 23-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Isaacson, John P.
; REGISTRATION NUMBER: 33,751
; REFERENCE/DOCKET NUMBER: 040433/0148
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-09-413-492-98

Query Match 35.7%; Score 5; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTL 13
Db 7 DGNTL 11

RESULT 13
US-08-726-136-7
; Sequence 7, Application US/08726136
; Patent No. 5811286
; GENERAL INFORMATION:
; APPLICANT: ROBERT D. FALLON
; APPLICANT: MARK S. PAYNE
; APPLICANT: MARK J. NELSON
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
; SOFTWARE: MICROSOFT WORD 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/004914
; FILING DATE: OCTOBER 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA A.
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
;
US-09-103-434-7

;
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
;
US-08-726-136-7

Query Match 35.7%; Score 5; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AWDVP 6
Db 9 AWDVP 13

RESULT 14
US-09-103-434-7
; Sequence 7, Application US/09103434
; Patent No. 6133421
; GENERAL INFORMATION:
; APPLICANT: ROBERT D. FALLON
; APPLICANT: MARK S. PAYNE
; APPLICANT: MARK J. NELSON
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
; SOFTWARE: MICROSOFT WORD 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,434
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,136
; FILING DATE:
; APPLICATION NUMBER: 60/004914
; FILING DATE: OCTOBER 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA A.
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
;
US-09-103-434-7
```

Query Match 35.7%; Score 5; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AWDVP 6
DB 9 AWDVP 13

RESULT 15
US-09-687-594-7
; Sequence 7, Application US/09687594
; Patent No. 6251650
; GENERAL INFORMATION:
; APPLICANT: ROBERT D. FALLON
; APPLICANT: MARK S. PAYNE
; APPLICANT: MARK J. NELSON
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
; SOFTWARE: MICROSOFT WORD 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/687,594
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/726,136
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA A.
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
US-09-687-594-7

Query Match 35.7%; Score 5; DB 4; Length 16;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AWDVP 6
DB 9 AWDVP 13

Search completed: January 31, 2002, 13:15:11
Job time: 92 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:09 ; Search time 78.64 Seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-73

Perfect score: 14
Sequence: 1 FAWVDPGWGNTLM 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	78.6	154	2 G75030	dCTP deaminase (EC
2	11	78.6	156	2 E71216	dCTP deaminase (EC
3	6	42.9	130	2 T31661	hypothetical prote
4	6	42.9	401	2 T48164	hypothetical prote
5	6	42.9	606	2 S77654	isocitrate lyase (
6	6	42.9	687	2 H72485	probable hydantoin
7	6	42.9	785	2 B72608	probable hyuA APE1
8	6	42.9	803	2 B82099	surface antigen VC
9	5	35.7	35	2 C44318	lactococcin G pept
10	5	35.7	115	2 E85592	hypothetical prote
11	5	35.7	115	2 B85883	probable transcrip
12	5	35.7	122	2 A82306	conserved hypothet
13	5	35.7	130	2 C72247	ribosomal protein
14	5	35.7	142	1 R5H025	ribosomal protein
15	5	35.7	142	2 S30800	ribosomal protein
16	5	35.7	142	2 S29999	ribosomal protein
17	5	35.7	153	2 G81236	smg protein, proba
18	5	35.7	169	2 E70540	hypothetical prote
19	5	35.7	171	2 A46618	platelet aggregati
20	5	35.7	177	2 H83695	deoxycytidine triph
21	5	35.7	178	2 S77760	probable ABC-type
22	5	35.7	179	2 A33104	tenebrosin C - sea
23	5	35.7	183	2 B24479	calcium-binding pr
24	5	35.7	185	2 S13183	calcium-binding pr
25	5	35.7	185	2 S13184	calcium-binding pr
26	5	35.7	185	2 S13185	calcium-binding pr
27	5	35.7	198	2 S54177	FSR protein - Ente
28	5	35.7	200	1 A42725	nitrile hydratase
29	5	35.7	207	2 S19716	nitrile hydratase

30 5 35.7 210 2 JH0603 beta-crystallin A4
31 5 35.7 214 2 JC4682 equinotoxin II pre
32 5 35.7 220 1 G69047 conserved hypothet
33 5 35.7 225 2 B35387 hypothetical prote
34 5 35.7 238 2 H70866 hypothetical prote
35 5 35.7 248 2 G75196 hypothetical prote
36 5 35.7 279 2 H84361 proline dehydrogen
37 5 35.7 280 2 T04049 chlorophyll a/b-bi
38 5 35.7 281 2 G86822 chlorophyll a/b-bi
39 5 35.7 283 2 T02250 chlorophyll a/b-bi
40 5 35.7 283 2 T02251 chlorophyll a/b-bi
41 5 35.7 286 2 S21386 chlorophyll a/b-bi
42 5 35.7 286 2 S16294 chlorophyll a/b-bi
43 5 35.7 289 2 H69645 myo-inositol catab
44 5 35.7 294 2 C82643 daunorubicin C-13
45 5 35.7 303 2 C69412 ABC transporter, A

ALIGNMENTS

RESULT 1
G75030
dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75030
R:anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: G75030
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50685.1; PID:g545
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: dcd; PAB1164
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 78.6%; Score 11; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FAWVDPGWGDN 11
|||||
Db 91 FAWVDPGWGDN 101

RESULT 2
E71216
dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: E71216
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Hailkawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilii
A:Reference number: A71000; MUID:98344137
A:Accession: E71216
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-156 <RAW>
A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31124.1; PID:g3258441
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH1997
C:Superfamily: dCTP deaminase

C;Keywords: hydrolase

Query Match 78.6%; Score 11; DB 2; Length 156;
Best Local Similarity 100.0%; Pred. No. 1;le-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FAWVDPGWDGN 11
|||||
Db 91 FAWVDPGWDGN 101

RESULT 3

T31661

hypothetical protein COS41.7 - sea squirt (Ciona intestinalis)

C;Species: Ciona intestinalis

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C;Accession: T31661

R;Bird, A.P.; Clark, V.; Jones, S.J.; Leitgeb, S.; Dobson, R.; Tweedie, S.

submitted to the EMBL Data Library, December 1996

A;Reference number: 221049

A;Accession: T31661

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-190 <BIR>

A;Cross-references: EMBL:Z83760; NID:81014349; PID:e289997; PIDN:CAB06051.1

C;Genetics:

A;Introns: 56/1; 87/1

Query Match 42.9%; Score 6; DB 2; Length 190;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 WVDPCW 8
|||||
Db 71 WVDPCW 76

RESULT 4

T48164

hypothetical protein T1008.i30 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C;Accession: T48164

R;Bavan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Lemcke, K.; Mayer, K.

submitted to the Protein Sequence Database, March 2000

A;Reference number: 224486

A;Accession: T48164

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-401 <BEV>

A;Cross-references: EMBL:AL161746

A;Experimental source: cultivar Columbia; BAC clone T1008

C;Genetics:

A;Map position: 5

A;Note: T1008.i30

Query Match 42.9%; Score 6; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTLM 14
|||||
Db 235 DGNTLM 240

RESULT 5

S77654

isocitrate lyase (EC 4.1.3.1) - Mycobacterium leprae

N;Alternate names: isocitrate; isocitratase; isocitritase

C;Species: Mycobacterium leprae

C;Date: 11-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 22-Jun-1999
C;Accession: S77654; S49515

R;Fishi, H.; Cole, S.T.

Mol. Microbiol. 16, 909-919, 1995

A;Title: The Mycobacterium leprae genome: systematic sequence analysis identifies key

A;Reference number: S77652; MUID:96059637

A;Accession: S77654

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-606 <FSI>

A;Cross-references: EMBL:Z46257; NID:9559905; PIDN:CAA86357.1; PID:9559906

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994

C;Genetics:

A;Gene: aceA

C;Superfamily: isocitrate lyase

C;Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 42.9%; Score 6; DB 2; Length 606;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VDPGWD 9
|||||
Db 413 VDPGWD 418

RESULT 6

H72485

probable hydantoinase APE2528 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 16-Feb-2001

C;Accession: H72485

R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A;Reference number: A72450; MUID:99310339

A;Accession: H72485

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-687 <KAW>

A;Cross-references: DDBJ:AP000064; NID:95105945; PIDN:BA81544.1; PID:95106233

A;Experimental source: strain K1

C;Genetics:

A;Gene: APE2528

C;Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyua

Query Match 42.9%; Score 6; DB 2; Length 687;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AWVDPG 7
|||||
Db 348 AWVDPG 353

RESULT 7

B72608

probable hyua APE1328 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 16-Feb-2001

C;Accession: B72608

R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.

DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A;Reference number: A72450; MUID:99310339

A;Accession: B72608

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-785 <KAW>

A;Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BAA80320.1; PID:g5105006
 A;Experimental source: strain K1
 C;Genetics:
 A;Gene: APE1328
 C;Superfamily: Pseudomonas D-amino acid hydantoin hydrolase (ATP-hydrolyzing) hyuA

Query Match 42.9%; Score 6; DB 2; Length 785;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AWDPG 7
 |||||
 Db 448 AWDPG 453

RESULT 8
 B82099
 surface antigen VC2252 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C;Species: Vibrio cholerae
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C;Accession: B82099
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A;Reference number: A82035; MUID:20406833
 A;Accession: B82099
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-803 <HEI>
 A;Cross-references: GB:AE004297; GB:AE003852; NID:g9656810; PIDN:AAF95396.1; GSPDB:GN001
 A;Experimental source: serogroup O1; strain N16961; biotype El Tor
 C;Genetics:
 A;Gene: VC2252
 A;Map position: 1
 C;Superfamily: protective surface antigen D-15

Query Match 42.9%; Score 6; DB 2; Length 803;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTLM 14
 |||||
 Db 80 DGNTLM 85

RESULT 9
 C44918
 lactococcin G peptide beta - Lactococcus lactis
 C;Species: Lactococcus lactis
 C;Date: 01-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 23-Mar-1995
 C;Accession: C44918
 R;Nissen-Meyer, J.; Holo, H.; Havarstein, L.S.; Sletten, K.; Nes, I.F.
 J. Bacteriol. 174, 5686-5692, 1992
 A;Title: A novel lactococcal bacteriocin whose activity depends on the complementary act
 A;Reference number: A44918; MUID:92380946
 A;Accession: C44918
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 1-35 <NIS>
 A;Experimental source: LMG 2081
 A;Note: sequence extracted from NCBI backbone (NCBIP:112782)

Query Match 35.7%; Score 5; DB 2; Length 35;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AWDPD 6
 |||||

Db 7 AWDPD 11
 RESULT 10
 E85592
 hypothetical protein 21051 [imported] - Escherichia coli (strain O157:H7)
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C;Accession: E85592
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: E85592
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-115 <STO>
 A;Cross-references: GB:AE005174; NID:g12513834; PIDN:AAG55201.1; GSPDB:GN00145; UWGP:
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: Z1051

Query Match 35.7%; Score 5; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
 |||||
 Db 88 DGNTL 92

RESULT 11
 B85883
 probable transcription regulator LYSR-type yfeR_2 [imported] - Escherichia coli (stra
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C;Accession: B85883
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: B85883
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-115 <STO>
 A;Cross-references: GB:AE005174; NID:g12516780; PIDN:AAG57526.1; GSPDB:GN00145; UWGP:
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: yfeR_2

Query Match 35.7%; Score 5; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GNTLM 14
 |||||
 Db 100 GNTLM 104

RESULT 12
 A82306
 conserved hypothetical protein VC0580 [imported] - Vibrio cholerae (strain N16961 ser
 C;Species: Vibrio cholerae
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C;Accession: A82306
 R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A:Reference number: A82035; MUID:20405833

A:Accession: A82306

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-122 <HEI>

A:Cross-references: GB:AE004143; GB:AE003852; NID:g9655003; PIDN:AAF93748.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0580

A:Map position: 1

C:Superfamily: hypothetical protein H1656

Query Match 35.7%; Score 5; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
|||||
Db 46 DGNTL 50

RESULT 13

C72247

ribosomal protein S11 - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: C72247

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Haft, D.H.; Hickey

Garratt, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; MUID:99287316

A:Accession: C72247

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-130 <ARN>

A:Cross-references: GB:AE001798; GB:AE000512; NID:g4982033; PIDN:AAD36542.1; PID:g498204

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1474

C:Superfamily: *Escherichia coli* ribosomal protein S11

Query Match 35.7%; Score 5; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
|||||
Db 39 DGNTL 43

RESULT 14

R5HQ25

ribosomal protein L23a.e - yeast (*Pichia jadinii*)

N:Alternate names: ribosomal protein HL25

C:Species: *Pichia jadinii*, *Candida utilis*

C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 22-Jun-1999

C:Accession: S06373

R:Woudt, L.P.; Mager, W.H.; Beek, J.G.; Wassenaar, G.M.; Planta, R.J.

Curr. Genet. 12, 193-198, 1987

A:Title: Structural and putative regulatory sequences of the gene encoding ribosomal pro

A:Reference number: S06373; MUID:88210534

A:Accession: S06373

A:Molecule type: DNA

A:Residues: 1-142 <WOU>

A:Cross-references: EMBL:X05919; NID:g2679; PIDN:CAA29354.1; PID:g2680

A:Note: the authors' translation is inconsistent with the nucleotide sequence in having

C:Comment: This protein binds to 26S rRNA.

C:Genetics:

A:Introns: 5/1
C:Superfamily: rat ribosomal protein L23a
C:Keywords: protein biosynthesis; ribosome; RNA binding

Query Match 35.7%; Score 5; DB 1; Length 142;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
|||||
Db 78 DGNTL 82

RESULT 15

S30000

ribosomal protein L23a.e - yeast (*Kluyveromyces marxianus*)

N:Alternate names: ribosomal protein YL25

C:Species: *Kluyveromyces marxianus*

C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 13-Aug-1999

C:Accession: S30000

R:Bergkamp-Steffens, G.K.; Hoekstra, R.; Planta, R.J.

Yeast 8, 903-922, 1992

A:Title: Structural and putative regulatory sequences of *Kluyveromyces* ribosomal prot

A:Reference number: S29999; MUID:93127726

A:Accession: S30000

A:Molecule type: DNA

A:Residues: 1-142 <BER>

A:Cross-references: GB:S53422; NID:g263476; PIDN:AAB24897.1; PID:g263477

A:Note: the authors did not show the third nucleotide in the codon for 10-Ala

C:Genetics:

A:Gene: L25

A:Introns: 5/1

C:Superfamily: rat ribosomal protein L23a

C:Keywords: protein biosynthesis; ribosome; RNA binding

Query Match 35.7%; Score 5; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
|||||
Db 78 DGNTL 82

Search completed: January 31, 2002, 13:20:11
Job time: 112 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:06 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-73

Perfect score: 14

Sequence: 1 FAWVDPGWGNTLM 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_1101.*

1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	19	AAW72849
2	11	78.6	156	19	AAW72847
3	6	42.9	55	20	AAW40052
4	6	42.9	106	22	AAU14180
5	6	42.9	561	18	AAW18311
6	6	42.9	891	20	AAW34495
7	6	42.9	899	20	AAW34369
8	5	35.7	9	22	AAU15304
9	5	35.7	13	15	AAW67105
10	5	35.7	13	21	AAW14944
11	5	35.7	16	18	AAW17823

12	5	35.7	27	16	AAW65038
13	5	35.7	27	19	AAW46641
14	5	35.7	27	20	AAW24499
15	5	35.7	41	22	AG73565
16	5	35.7	43	22	AG75841
17	5	35.7	54	20	AAW00340
18	5	35.7	54	20	AAW00273
19	5	35.7	55	21	AAW33744
20	5	35.7	58	22	AAW18941
21	5	35.7	58	22	AAW31515
22	5	35.7	60	15	AAW49133
23	5	35.7	74	17	AAW95262
24	5	35.7	94	22	AAW18388
25	5	35.7	94	22	AAW30866
26	5	35.7	94	22	AAW05984
27	5	35.7	121	22	AG74048
28	5	35.7	123	22	AG72675
29	5	35.7	131	20	AAW78444
30	5	35.7	135	22	AG92317
31	5	35.7	142	22	AG70915
32	5	35.7	143	19	AAW59862
33	5	35.7	148	21	AAW42848
34	5	35.7	148	21	AAW14942
35	5	35.7	159	21	AG26215
36	5	35.7	159	21	AG42746
37	5	35.7	159	22	AG76100
38	5	35.7	164	20	AAV19895
39	5	35.7	171	14	AAW41488
40	5	35.7	172	19	AAW59860
41	5	35.7	175	19	AAW59857
42	5	35.7	193	17	AAW01616
43	5	35.7	200	12	AAW13831
44	5	35.7	207	12	AAW13902
45	5	35.7	210	18	AAW17816

ALIGNMENTS

RESULT 1
AAW72849 standard; Peptide; 14 AA.
XX
AC AAW72849;
XX
DT 01-MAR-1999 (first entry)
XX
DE Pyrococcus furiosus p45 dUTPase uridine-binding motif.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX
(STRA-) STRATAGENE.
XX
Hansen CJ, Hogrefe H;
XX
WPI; 1998-542284/46.
XX
PT Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication
XX

PS Claim 71; Page 47; 161pp; English.

XX This is the uridine-binding motif of the P45 component (see
CC AAW72847) of the polymerase enhancing factor (PEF) of *Pyrococcus*
CC *furiosus* DSM 3638. P45 has been identified as a dUTPase that
CC has polymerase enhancing activity. Sequences are provided (see
CC AAW72849-57) of the uridine-binding motifs of dUTPases and dCYP
CC deaminases of *P. furiosus*, *Methanococcus jannaschii*, *Desulfurolobus*
CC *ambivalens*, *Escherichia coli* yeast, human and herpesvirus A
CC consensus (see AAW72848) is also provided. A claimed method of
CC enhancing a nucleic acid polymerase reaction comprises performing
CC the reaction in the presence of one or more of the following: a
CC PEF, a dUTPase, a PEF that turns-over dUTP and a protein
CC having one or more of the sequences provided in AAW72848-57. A
CC claimed protein having PEF activity comprises one or more of
CC sequences given in AAW72848-57. Kits are provided for replicating
CC nucleic acids, for site-directed mutagenesis, for nucleic acid
CC sequencing or for amplification (preferably PCR or RT-PCR).

XX Sequence 14 AA;

Query Match 100.0%; Score 14; DB 19; Length 14;
Best Local Similarity 100.0%; Pred. No. 6e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAWVDPGWDGNTLM 14
Db 1 fawvdpwgwdgntlm 14

RESULT 2

ID AAW72847
AC AAW72847

DT 01-MAR-1999 (first entry)

DE Polymerase enhancing factor P45 (dUTPase) component.

KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication.

OS *Pyrococcus furiosus* strain DSM 3638.

FH Key Location/Qualifiers

FT Peptide 2..15

FT /note= "N-terminal peptide used to generate
FT primers"

XX WO9842860-A1.

XX 01-OCT-1998.

XX 20-MAR-1998; 98WO-US05497.

XX 24-OCT-1997; 97US-0957709.

XX 21-MAR-1997; 97US-0822774.

XX (STRA-) STRATAGENE.

XX Hansen CJ, Hogrefe H;

XX WPI: 1998-542284/46.

XX N-PSDB; AAW63860.

XX Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication

PS Claim 17; Page 43; 161pp; English.

CC This is the amino acid sequence of the P45 component of the
CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.
CC The sequence is predicted from a DNA sequence, see AAW63860) obtained
CC from genomic DNA by PCR. P45 and P50 (see AAW72844) are the activity of
CC frequent components of PEF which acts to enhance the activity of
CC *Pyrococcus* DNA polymerase. P45 functions as a dUTPase, and can be
CC used to enhance nucleic acid replication, polymerisation, or PCR
CC reactions. The invention provides novel extracts, proteins and
CC complexes that improve the polymerisation activity of nucleic acid
CC polymerases, as well as DNA constructs and antibodies. Also
CC included are methods for identifying compositions with PEF activity,
CC for purifying and using these compositions, and specific extracts,
CC proteins and complexes that function to enhance polymerase activity.
CC Nucleic acid polymerase reactions can be enhanced (claimed) by
CC mixing a nucleic acid template, at least 1 polymerase and a
CC composition having polymerase enhancing activity. Kits are
CC provided for replicating nucleic acids, for site-directed
CC mutagenesis, for nucleic acid sequencing or amplification
CC (preferably PCR or RT-PCR).

XX Sequence 156 AA;

Query Match 78.6%; Score 11; DB 19; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FAWVDPGWDGDN 11
Db 91 fawvdpwgwdgn 101

RESULT 3

AAAY40052

ID AAY40052 standard; Peptide: 55 AA.

XX AC AAY40052;

XX 18-NOV-1999 (first entry)

DE Peptide sequence derived from a human secreted protein.

XX Secreted protein; gene therapy; cancer; tumor; fetal deficiency;
KW neurodegenerative disorder; developmental abnormality; blood disorder;
KW immune system disease; autoimmune disease; leukemia; inflammation;
KW allergy; Alzheimer's disease; cognitive disorder; schizophrenia;
KW obesity; osteoporosis; arthritis; infection; AIDS; diabetes; asthma;
KW connective tissue disorder; transplant rejection; sepsis; acne;
KW psoriasis; cardiovascular disorder; reproductive disorder;
KW food additive; food preservative; storage capability.

XX Homo sapiens.

XX WO9943693-A1.

XX 02-SEP-1999.

XX 24-FEB-1999; 99WO-US03939.

XX 26-FEB-1998; 98US-0076051.

XX 26-FEB-1998; 98US-0076052.

XX 26-FEB-1998; 98US-0076053.

XX 26-FEB-1998; 98US-0076054.

XX 26-FEB-1998; 98US-0076057.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Olsen HS, Florence K, Brewer LA, Ebner R, Ruben SM, Rosen CA;

XX Duan RD;

XX WPI: 1999-550857/46.

XX New human genes and the secreted polypeptides they encode, useful for

PT diagnosis and treatment of e.g. cancers, neurological disorders, immune
 PT diseases, inflammation or blood disorders
 XX
 PS Disclosure; Page 44; 246pp; English.
 XX
 CC AAY40001-92 are derived from human secreted proteins. The
 CC polynucleotides and their corresponding secreted polypeptides are useful
 CC for preventing, treating or ameliorating medical conditions, e.g. by
 CC protein or gene therapy. Pathological conditions can also be diagnosed by
 CC determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the polynucleotide. Specific
 CC uses include developing products for the diagnosis or treatment of
 CC cancer, tumors, neurodegenerative disorders, developmental abnormalities
 CC and fetal deficiencies, blood disorders, sepsis, diseases of the immune
 CC system, autoimmune diseases, inflammation, allergies, Alzheimer's and
 CC cognitive disorders, schizophrenia, obesity, osteoporosis, arthritis,
 CC infections, AIDS, connective tissue disorders, transplant rejection,
 CC diabetes, asthma, sepsis, acne, psoriasis, cardiovascular disorders,
 CC and reproductive disorders. The polypeptides or polynucleotides can
 CC also be used as food additives or preservatives, such as to increase
 CC or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors or other nutritional
 CC components.
 XX
 SQ Sequence 55 AA;

Query Match 42.9%; Score 6; DB 20; Length 55;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AWDVDPG 7
 Db 39 awdvpq 44
 RESULT 4
 AAW14180
 ID AAW14180 standard; Protein; 106 AA.
 XX
 AC AAW14180;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human novel protein #51.
 XX
 KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; Immune disorder.
 XX
 OS Homo sapiens..
 XX
 PN WO200155437-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001WO-US02623.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-451939/48.
 DR N-PSDB; AAS22485.
 XX
 PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -

XX
 PS Example 4; Page 556; 894pp; English.
 XX
 CC The invention relates to polynucleotides encoding novel human
 CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/elicit an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.
 XX
 SQ Sequence 106 AA;

Query Match 42.9%; Score 6; DB 22; Length 106;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AWDVDPG 7
 Db 91 awdvpq 96
 RESULT 5
 AAW18311
 ID AAW18311 standard; Protein; 561 AA.
 XX
 AC AAW18311;
 XX
 DT 10-AUG-1997 (first entry)
 XX
 DE Rabbit complement C9.
 XX
 KW Complement C9; CD59; C5b9 complex; tumour; therapy; contraceptive;
 KW antibody; inflammation..
 XX
 OS Oryctolagus cuniculus.
 XX
 FH Key Location/Qualifiers
 FT Disulfide-bond 389..418
 XX
 PN WO9717987-A1.
 XX
 PD 22-MAY-1997.
 XX
 XX 08-NOV-1996; 96WO-US17940.
 XX
 XX 15-NOV-1995; 95US-0559492.
 XX
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.
 XX
 PI Sims PJ;
 XX
 DR WPI; 1997-289058/26.
 DR N-PSDB; T688887.

XX Modulating binding of CD59 to C9 complement component - uses agent
 PT that mimics or binds the C9-specific motif, used to activate
 PT complement in tumour therapy or to treat complement mediated
 PT inflammation
 PS Disclosure; Page 36-38; 51pp; English.
 XX
 CC A polypeptide (AAW19311) comprises rabbit complement C9. Chimeric
 CC proteins were constructed in which the segment of C9 corresponding
 CC to the putative CD59 binding site were interchanged between rabbit
 CC and human C9 (see also AAW18310). These were tested for haemolytic
 CC activity and for sensitivity to inhibition by membrane CD59.
 CC Substn. of human C9 residues 334-415 into rabbit C9 resulted in a
 CC protein that was indistinguishable from human C9 in its sensitivity
 CC to inhibition by CD59. When this same segment of human C9 was
 CC replaced by the corresponding rabbit sequence (aa339-424), the
 CC chimera was indistinguishable from rabbit C9 and was virtually
 CC unaffected by the presence of membrane CD59.
 XX
 SQ Sequence 561 AA;

Query Match 42.9%; Score 6; DB 18; Length 561;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 WDGNPL 13
 Db 190 wdgnl 195
 |||||

RESULT 6
 AAY34495
 ID AAY34495 standard; Protein: 891 AA.
 XX
 AC AAY34495;
 XX
 DT 25-AUG-1999 (first entry)
 XX
 DE Porphyromonas gingivalis protein PG36.
 XX
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX
 OS Porphyromonas gingivalis.

XX WO9929870-A1.
 PN 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98WO-AU01023.
 XX
 PR 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.

(CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 XX Ross BC, Rothel LJ, Webb EA;
 XX WPI; 1999-385613/32.
 DR N-PSDB; AAX91713.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing

PT gingivitis
 XX
 PS Claim 1; Page 476-477; 588pp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 XX
 SQ Sequence 891 AA;

Query Match 42.9%; Score 6; DB 20; Length 891;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FAWVDP 6
 Db 359 fawvdp 364
 |||||

RESULT 7
 AAY34369
 ID AAY34369 standard; Protein: 899 AA.
 XX
 AC AAY34369;
 XX
 DT 25-AUG-1999 (first entry)
 XX
 DE Porphyromonas gingivalis protein PG36.
 XX
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
 KW vaccine; antigenic.
 XX
 OS Porphyromonas gingivalis.

XX WO9929870-A1.
 PN 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98WO-AU01023.
 XX
 PR 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.

(CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 XX Ross BC, Rothel LJ, Webb EA;
 XX WPI; 1999-385613/32.
 DR N-PSDB; AAX91587.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX
 PS Claim 1; Page 332-333; 588pp; English.
 XX
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to

CC AAY34583. AAX91802 to AAX91889 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 CC
 XX Sequence 899 AA;

SQ Sequence 899 AA;

Query Match 42.9%; Score 6; DB 20; Length 899;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FAWYDP 6
 Db 367 fawvdp 372

RESULT 8
 AAR67105
 ID AAR67105 standard; Peptide; 9 AA.

XX AC AAR67105;
 XX XX 30-JUN-1995 (first entry)

DE Schizophrenia-associated isoform peptide #189.

XX Schizophrenia; neuroleptic; diagnostic; neuropsychiatric disorder;
 KW neurological disorder; neuropathy.

XX Homo sapiens.

XX WO200163293-A2.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-GB00783.

XX 24-FEB-2000; 2000GB-0004415.

PR 28-NOV-2000; 2000US-0750395.

XX (OXFO-) OXFORD GLYSCSCIENCES UK LTD.

XX Herath HMAC, Parekh RB, Rohlf C;

XX WPI; 2001-502868/55.

XX Diagnosing and monitoring Schizophrenia by detecting the presence of
 PT Schizophrenia Associated Features and Schizophrenia Associated Protein
 PT Isoforms in samples of cerebrospinal fluid -

XX Claim 6; Page 32; 160pp; English.

XX The invention relates to methods and compositions for screening,
 CC diagnosis and prognosis of Schizophrenia. The method involves detecting
 CC the presence of Schizophrenia (SCH) Associated Features (Sf) and SCH
 CC Associated Protein Isoforms (SPIs) in samples, e.g. by electrophoresis,
 CC immunassay or hybridisation assay, for diagnosing and monitoring SCH,
 CC studying the effectiveness of treatments and for identifying potential
 CC therapeutic agents. The method is used for (1) screening or diagnosis of
 CC SCH and the relative abundance of at least 1 chosen feature correlates
 CC with the presence or absence of SCH; and (2) monitoring the effect of
 CC therapy administered to a subject with SCH and the relative abundance of
 CC at least 1 chosen feature which correlates with the severity of SCH.
 CC The expression and activity of the Sf, SPIs and related molecules
 CC (e.g. secondary messengers) are studied to diagnose SCH, monitor the
 CC progress of the disorder and the effectiveness of treatment and as
 CC targets to identify and produce potential therapeutic agents for the
 CC treatment of SCH. The paucity of detectable neuralgic defects
 CC distinguishes neuropsychiatric disorders such as SCH from neurological

CC disorders, where manifestations of anatomical and biochemical changes
 CC have been identified in many cases. Consequently the identification and
 CC characterisation of cellular and/or molecular causative defects and
 CC neuropathies are necessary for improved treatment of neuropsychiatric
 CC disorders. AAR67105 represents the amino acid sequences of
 CC schizophrenia-associated isoforms used in the method of the invention.
 XX
 SQ Sequence 9 AA;

Query Match 35.7%; Score 5; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTL 13
 Db 1 dgntl 5

RESULT 9.

AAR67105
 ID AAR67105 standard; peptide; 13 AA.

XX AC AAR67105;

XX XX 30-JUN-1995 (first entry)

DE Antiplatelet peptide (APCol-M2) contg. proline brackets.

XX antiplatelet peptide; leech antiplatelet protein; collagen receptor;
 KW mimic; interaction site; constrained conformation; aggregation; moubatin.

XX Synthetic.

XX WO9425482-A.

XX 10-NOV-1994.

XX 21-APR-1994; 94WO-US04294.

XX 23-APR-1993; 93US-0051741.

PR 29-OCT-1993; 93US-0143364.

XX (EVAN/) EVANS H J.

PA (KINI/) KINI R M.

XX Evans HJ, Kini RM;

XX WPI; 1994-358186/44.

XX Peptide homologue or analogue with constrained conformation - has
 PT proline residues flanking the interaction site to impart greater,
 PT or more stable, biological activity

XX Example 2; Page 36; 57pp; English.

XX AAR67102-5 are antiplatelet peptides derived from naturally occurring
 CC polypeptides that contain proline or proline/cysteine brackets. These
 CC peptides are shortened to form fragments that contain one or more
 CC interaction sites of interest. AAR67104-105 are deriv. from moubatin.
 CC The dose is 1-4 micromoles. The peptides interact with specific
 CC receptors and inhibit tumour growth. The data collected demonstrates
 CC that interaction sites possess activity when present in a polypeptide
 CC that differs from the native form. Inclusion of
 CC conformation-constraining moieties can have desirable effects on an
 CC interaction site. (Also see AAR67011-101 and AAR67106-52 for analogues of
 CC other biologically active peptides contg. an interaction site flanked by
 CC conformation constraining gps., eg. RGD peptides.)

XX Sequence 13 AA;

Query Match 35.7%; Score 5; DB 15; Length 13;

Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNL 13
Db 7 dgentl 11

RESULT 10

AA14944
ID AAB14944 standard; peptide; 13 AA.

XX AC AAB14944;

XX DT 09-JAN-2001 (first entry)

XX DE Porphyromonas gingivalis Prtr27 peptide #2.

XX KW Porphyromonas gingivalis; protective epitope; antiinflammatory; vaccine;
XX KW periodontal disease; periodontitis; immunogen; infection; Prtr27 adhesin.

XX OS Porphyromonas gingivalis.

XX PN WO200052041-A1.

XX PD 08-SEP-2000.

XX PF 01-MAR-2000; 2000WO-AU00142.

XX PR 01-MAR-1999; 99AU-0008939.

XX PA (UYME) UNIV MELBOURNE.

XX PA (CSLC-) CSL LTD.

XX PA (VICT-) VICTORIAN DAIRY IND ASSOC.

XX PI O'Brien-Simpson NM, Reynolds EC;

XX DR WPI; 2000-579275/54.

XX PT New compositions comprising at least one peptide with an epitope of
XX PT Arg-specific proteinase associated with Lys-specific proteinase protein
XX PT complex for treating and preventing periodontitis-associated
XX PT Porphyromonas gingivalis

XX PS Example 1; Page 21; 62pp; English.

XX CC The present sequence is one of twenty overlapping peptides corresponding
XX CC to the N-terminal 148 residues of the Prtr27 adhesin from Porphyromonas
XX CC gingivalis. Antibodies directed towards the Prtr27 adhesin were found to
XX CC provide protection against periodontitis in a human patient and
XX CC immunoprotected mice. The peptides were used for epitope mapping of the
XX CC Prtr27 adhesin. Peptides comprising at least one P. gingivalis epitope
XX CC are useful in the treatment and prevention of periodontal disease,
XX CC particularly periodontitis, which is associated with P. gingivalis
XX CC infection. The peptides are useful as immunogens in vaccine formulations
XX CC for active immunisation. They may also be used to generate
XX CC protein-specific and peptide-specific antiserum useful for passive
XX CC immunisation, and to generate polyclonal or monoclonal antibodies. The
XX CC peptides and antibodies against the peptides can be used in diagnostic
XX CC tests to detect P. gingivalis infection. The antibodies can also be used
XX CC in oral compositions such as toothpaste and mouthwash to neutralise the
XX CC antigen and prevent disease.

XX SQ Sequence 13 AA;

Query Match 35.7%; Score 5; DB 21; Length 13;

Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 WDGNT 12
Db 7 wdgt 11

RESULT 11

AAW17823
ID AAW17823 standard; Peptide; 16 AA.

XX AC AAW17823;

XX DT 29-JUL-1997 (first entry)

XX DE Nitrile hydratase alpha subunit partial sequence.

XX KW Nitrile hydratase; stereospecific enzyme; chiral compound; amide.

XX OS Pseudomonas putida strain NRRL 18668.

XX PN WO9712964-A2.

XX PD 10-APR-1997.

XX PF 03-OCT-1996; 96WO-US15969.

XX PR 06-OCT-1995; 95US-0004914.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PI Fallon RD, Nelson MJ, Payne MS;

XX DR WPI; 1997-226208/20.

XX CC Nucleic acid encoding stereospecific nitrile hydratase and amidase -
XX CC also transformed cells expressing these enzymes; useful for
XX CC conversion of racemic nitrile compounds to chiral amide(s) and acids
XX CC Example 1; Page 49; 85pp; English.

XX CC Peptides (AAW17821-25) were obtd. by subjecting the purified
XX CC nitrile hydratase alpha subunit (AAW17816) of Pseudomonas putida
XX CC NRRL 18668 to cyanogen bromide, IPCK-treated trypsin and AspN
XX CC protease digestions. Peptides (AAW17826-29) of the beta subunit
XX CC (AAW17817) were also obtd. PCR primers (AAW6643-44) based on the
XX CC peptides were used to amplify 18668 genomic DNA. A partial
XX CC nitrile hydratase sequence (AAW6645) was obtd. and used as a probe
XX CC to identify the full-length coding sequence (AAW66440).

XX SQ Sequence 16 AA;

Query Match 35.7%; Score 5; DB 18; Length 16;

Best Local Similarity 100.0%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AWDVP 6
Db 9 awvdp 13

RESULT 12

AAAR65038
ID AAR65038 standard; Peptide; 27 AA.

XX AC AAR65038;

XX DT 06-OCT-1995 (first entry)

XX DE Random biotinylation peptide 7.

XX KW biotinylation; peptide; recombinant; fusion protein; small;
XX KW specific; defined; purification; BirA; enzyme; biotin.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 16 /note= "biotin-Lys"
 FT XX WO9504069-A.
 PN XX
 PD 09-FEB-1995.
 XX
 XX 28-JUL-1994; 94WO-US08528.
 XX
 XX 30-JUL-1993; 93US-0099991.
 XX
 XX (AFFY-) AFFYMAX TECHNOLOGIES NV.
 XX
 XX Schatz PJ;
 XX WPI; 1995-090609/12.
 XX
 XX Production of biotinylated proteins by expression of a
 PT recombinant DNA vector - which encodes a fusion protein
 PT comprising a protein and a biotinylated peptide.
 XX
 XX Claim 10; Page 136; 146pp; English.
 XX
 CC A library of small, efficient peptide biotinylation sequences
 CC (AAR65032-46) was generated by using a generic peptide (AAR65020) and a
 CC system known as the "peptides on plasmids" system. At some positions in
 CC the sequences, no clear consensus is apparent. At other residues,
 CC however, clear trends emerge. A protein can be biotinylated by
 CC constructing a recombinant DNA expression vector encoding a fusion
 CC protein, comprising a protein and a biotinylation peptide. A host cell,
 CC eg. E. coli is transformed with the vector and is cultured in the
 CC presence of biotin and a biotinylation enzyme, eg. BirA.
 XX
 XX Sequence 27 AA;

Query Match 35.7%; Score 5; DB 16; Length 27;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGWDG 10
 Db |
 6 pgwdg 10

RESULT 13
 AAW46641
 ID AAW46641 standard; peptide; 27 AA.
 XX
 AC AAW46641;
 XX
 DT 28-MAY-1998 (first entry)
 XX
 DE Biotinylation peptide isolated from random library 1.
 XX
 XX Biotinylation peptide; biotinylation enzyme; biotin-protein ligase;
 KW BirA; biotin ligase; biotin; purification; immobilisation; labelling;
 KW detection; protein.
 XX
 OS Synthetic.
 XX
 XX US5723584-A.
 PN
 PD 03-MAR-1998.
 XX
 XX 03-FEB-1995; 95US-0383753.
 XX
 XX 30-JUL-1993; 93US-0099991.
 XX
 XX (AFFY-) AFFYMAX TECHNOLOGIES NV.
 PA
 XX Schatz PJ;
 XX

DR WPI; 1998-178542/16.
 XX
 XX Peptide(s) that can be biotinylated by biotin ligase - and fusion
 PT proteins containing them
 PT
 XX
 PS Claim 3; Column 65; 33pp; English.
 XX
 XX Peptides AAW46635-46 are non-naturally occurring biotinylation peptides,
 CC derived from a library constructed to express peptides of the generic
 CC sequence AAW46623. The library was constructed using oligonucleotides
 CC AAV16118-20. The peptides contain a biotinylatable sequence motif,
 CC recognised by a biotinylation enzyme, e.g. biotin-protein ligase (BirA).
 CC The C or N terminus of the peptides can be covalently coupled to a
 CC protein that is incapable of being biotinylated by a biotin ligase.
 CC The peptides can be biotinylated in vitro or in vivo, especially with
 CC BirA biotin ligase, and used for the purification, immobilisation,
 CC labelling or detection of proteins.
 XX
 XX Sequence 27 AA;

Query Match 35.7%; Score 5; DB 19; Length 27;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGWDG 10
 Db |
 6 pgwdg 10

RESULT 14
 AAY24499
 ID AAY24499 standard; peptide; 27 AA.
 XX
 AC AAY24499;

XX
 XX 28-SEP-1999 (first entry)
 XX
 DE Biotinylation peptide SEQ ID NO:20.
 XX
 XX Biotinylation enzyme; biotin-protein ligase; BirA; labelling;
 KW biotin carboxyl carrier protein; BCCP; Escherichia coli;
 KW fusion protein; identification; purification; diagnosis; research.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN US5932433-A.
 XX
 PD 03-AUG-1999.
 XX
 XX 28-OCT-1997; 97US-0959512.
 XX
 XX 03-FEB-1995; 95US-0383753.
 PR 30-JUL-1993; 93US-0099991.
 PR 28-OCT-1997; 97US-0959512.
 XX
 XX (AFFY-) AFFYMAX TECHNOLOGIES NV.
 PA
 XX Schatz PJ;
 XX
 XX WPI; 1999-457113/38.
 XX
 PT Identification and biotinylation of proteins synthesized by
 PT recombinant DNA techniques in vivo
 XX
 XX Claim 6; Column 65; 35pp; English.
 XX
 CC The present invention describes a method for the identification and
 CC biotinylation of proteins synthesized by recombinant DNA techniques
 CC in vivo with a biotinylation peptide of less than 50 amino acids. The
 CC method comprises: (a) on the surface of a substrate, providing a fusion
 CC protein comprising a recombinant protein and a peptide defined by the

CC sequence given in AAY24492, where the peptide is capable of being
 CC biotinylated by a biotin ligase at the lysine residue adjacent to
 CC position 8 and is 13-50 aa in length; (b) in a predefined region of the
 CC surface of the substrate, contacting the fusion protein with an enzyme;
 CC and (c) determining whether the fusion protein has been biotinylated.
 CC The method allows the identification and purification of biotinylated
 CC enzymes e.g. BirA. The method is also useful in research and diagnostic
 CC applications. The method uses small but specific peptides, allowing the
 CC labelling of a protein at a defined site, this provides improved
 CC immobilization and avoids the use of antibodies. AAY24493 to AAY24548,
 CC and AAY29299 to AAY29312 represent specifically claimed examples of
 CC biotinylation peptides for use in the method of the invention.

XX SQ Sequence 27 AA;

Query Match 35.7%; Score 5; DB 20; Length 27;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGWDG 10
 Db 6 pgwdg 10
 |||||

RESULT 15
 AAG73565
 ID AAG73565 standard; Protein; 41 AA.
 AC AAG73565;
 XX
 XX 03-SEP-2001 (first entry)
 DT Human colon cancer antigen protein SEQ ID NO:4329.
 XX
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 DE colorectal carcinoma.
 KW Homo sapiens.
 KW WO200122920-A2.
 XX
 XX 05-APR-2001.
 PD
 XX 28-SEP-2000; 2000WO-US26524.
 PF
 XX 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 DR
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI
 XX WPI: 2001-235357/24.
 DR N-PSDB; AAG73596.
 XX
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PT
 XX Claim 11; Page 6175-6176; 9803pp; English.
 PS
 XX AAG732943 to AAG737195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing P.
 CC Inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAG737196 to AAG737204
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 41 AA;

Query Match 35.7%; Score 5; DB 22; Length 41;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 DGNTL 13
 Db 20 dgntl 24
 |||||

Search completed: January 31, 2002, 13:18:07
 Job time: 168 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:38 ; Search time 130.99 Seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-73

Perfect score: 14
Sequence: 1 FAWDPGWGNTLM 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_17: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	42.9	190	5 P91585	P91585 ciona intes
2	6	42.9	401	10 Q9M031	Q9M031 arabidopsis
3	6	42.9	412	10 Q9LVP4	Q9LVP4 arabidopsis
4	6	42.9	687	1 Q9Y8V6	Q9Y8V6 aeropyrum p
5	6	42.9	785	1 Q9YCC8	Q9YCC8 aeropyrum p
6	6	42.9	803	2 Q9KPM0	Q9KPM0 vibrio chol
7	5	35.7	21	2 Q9R4L4	Q9R4L4 helicobacte
8	5	35.7	51	5 Q27101	Q27101 trichomonas
9	5	35.7	64	10 Q42130	Q42130 arabidopsis
10	5	35.7	92	2 Q45127	Q45127 bacillus fi
11	5	35.7	93	6 Q62727	Q62727 canis fami
12	5	35.7	104	10 P93074	P93074 bambusa sp.
13	5	35.7	116	5 Q27100	Q27100 trichomonas
14	5	35.7	122	2 Q9KUE1	Q9KUE1 vibrio chol
15	5	35.7	131	2 Q9A8K7	Q9A8K7 caulobacter
16	5	35.7	147	2 Q9AMN4	Q9AMN4 clostridium
17	5	35.7	153	2 Q9JXR28	Q9JXR28 neisseria m
18	5	35.7	169	2 Q06619	Q06619 mycobacteri
19	5	35.7	171	2 Q86213	Q86213 enterococcu

20	5	35.7	171	13 Q9PUF8	Q9PUF8 xenopus lae
21	5	35.7	177	2 Q9KRV3	Q9KRV3 bacillus ha
22	5	35.7	178	2 Q49088	Q49088 mycoplasma
23	5	35.7	190	2 Q9X2X0	Q9X2X0 rhodospiril
24	5	35.7	193	2 Q9AQF5	Q9AQF5 clostridium
25	5	35.7	198	2 Q47781	Q47781 enterococcu
26	5	35.7	199	2 Q9RJT7	Q9RJT7 streptomyce
27	5	35.7	220	1 Q26239	Q26239 methanobact
28	5	35.7	223	4 Q9H5T6	Q9H5T6 homo sapien
29	5	35.7	229	9 Q9X2X8	Q9X2X8 mycobacteri
30	5	35.7	234	2 Q9Z3W2	Q9Z3W2 pseudomonas
31	5	35.7	235	2 Q9AJY9	Q9AJY9 streptomyce
32	5	35.7	238	2 Q53200	Q53200 mycobacteri
33	5	35.7	239	2 Q9R704	Q9R704 agrobacteri
34	5	35.7	246	11 Q9CVB0	Q9CVB0 mus musculu
35	5	35.7	248	1 Q9V2H6	Q9V2H6 pyrococcus
36	5	35.7	248	2 Q9X749	Q9X749 helicobacte
37	5	35.7	253	2 Q9X4R1	Q9X4R1 enterococcu
38	5	35.7	265	13 Q9W682	Q9W682 oncorhynchu
39	5	35.7	265	13 Q9W680	Q9W680 oncorhynchu
40	5	35.7	273	5 Q9U9P7	Q9U9P7 drosophila
41	5	35.7	279	1 Q9HNG0	Q9HNG0 halobacteri
42	5	35.7	280	10 Q9XF89	Q9XF89 arabidopsis
43	5	35.7	281	2 Q9CF98	Q9CF98 lactococcus
44	5	35.7	283	10 Q41746	Q41746 zea mays (m
45	5	35.7	283	10 Q41747	Q41747 zea mays (m

ALIGNMENTS

RESULT 1
P91585 ID P91585 PRELIMINARY; PRT; 190 AA.
AC P91585;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COS41.7
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OC Clonidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird A.P., Clark V., Jones S.J.M., Leitch S., Dobson R., Tweedie S.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z83760; CAB06051.1; .
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR000387; TYR_phosphatase.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
DR PROSITE; PSS00056; TYR_PHOSPHATASE_2; 1.
DR SMART; SM00012; PTPc_DSPC; 1.
SQ SEQUENCE 190 AA; 21292 MW; F8934267DADAD27E CRC64;

Query Match 42.9%; Score 6; DB 5; Length 190;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 WVDPGW 8
| | | | |
Db 71 WVDPGW 76

RESULT 2
Q9M031 ID Q9M031 PRELIMINARY; PRT; 401 AA.
AC Q9M031;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 45.3 KDA PROTEIN.

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AWDPG 7
|||||

Db 448 AWDPG 453

RESULT 6

Q9KPMO Q9KPMO PRELIMINARY; PRT; 803 AA.
AC Q9KPMO;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SURFACE ANTIGEN.
GN VC2252.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RP "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL: AE004297; AAF95396.1; -
DR TIGR: VC2252; -
DR InterPro: IPR000184; Bac_surfact_D15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89930 MW; 1998B3838F6C041D CRC64;

Query Match 42.9%; Score 6; DB 2; Length 803;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTLM 14
|||||

Db 80 DGNTLM 85

RESULT 7

Q9R4L4 Q9R4L4 PRELIMINARY; PRT; 21 AA.
AC Q9R4L4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE PRUVATE-FLAVODOXIN OXIDOREDUCTASE (EC 1.2.7.1) (FRAGMENT).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE.
RX MEDLINE=95332198; PubMed=7608066;
RA Hughes N.J., Chalk P.A., Clayton C.L., Kelly D.J.;
RT "Identification of carboxylation enzymes and characterization of a
RT novel four-subunit pyruvate:flavodoxin oxidoreductase from
RT Helicobacter pylori";
RL J. Bacteriol. 177:3953-3959(1995).
SQ SEQUENCE 21 AA; 2319 MW; 127528027BAE8143 CRC64;

Query Match 35.7%; Score 5; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 WDNT 12
|||||

Db 12 WDNT 16

RESULT 8

Q27101 Q27101 PRELIMINARY; PRT; 51 AA.
AC Q27101;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE UBIQUITIN (FRAGMENT).
GN UBQC.
OS Trichomonas vaginalis.
OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonas.
OX NCBI_TaxID=5722;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIH-C1, ATCC 30001;
RX MEDLINE=96081479; PubMed=7490769;
RA Keeling P.J., Doolittle W.F.;
RT "Concerted evolution in protists: recent homogenization of a
RT polyubiquitin gene in Trichomonas vaginalis";
RL J. Mol. Evol. 41:556-562(1995).
DR EMBL: U28013; AAC46941.1; -
DR HSSP: P02248; IUBI.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 2.
DR PROSITE: PS50053; UBIQUITIN_2; 1.
FT NON_TER 1
FT NON_TER 51
SQ SEQUENCE 51 AA; 5810 MW; 947D7ED017FFF26C CRC64;

Query Match 35.7%; Score 5; DB 5; Length 51;

Best Local Similarity 100.0%; Pred. No. 40;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTL 13
|||||

Db 15 DGNTL 19

RESULT 9

Q42130 Q42130 PRELIMINARY; PRT; 64 AA.
AC Q42130;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GUANINE NUCLEOTIDE-BINDING PROTEIN (FRAGMENT).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-WHOLE SEEDLINGS OF A. THALIANA ECOTYPE COLUMBIA;
RA Hofte H.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z26552; CAA81323.1; -
DR Mendel: 16252; Arath:1350.16252.
DR InterPro: IPR001680; WD40.
DR SMART: SM00320; WD40; 1.
DR PROSITE: PS50082; WD_REPEATS_2; 1.
DR PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.

FT NON_TER 1 1
 SQ SEQUENCE 64 AA; 6945 MW; 5735E41922E70993 CRC64;

Query Match 35.7%; Score 5; DB 10; Length 64;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 DGNTL 13
 Db 43 DGNTL 47

RESULT 10
 Q45127 PRELIMINARY; PRT; 92 AA.
 AC 045127;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE SIMILAR TO VIBRIO PARAHAEVOLYTICUS MOTX PROTEIN.
 OS Bacillus firmus
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OX Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0F4;
 RA Ito M., Krulwich T.A.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U33209; AAA75478.1; -;
 SQ SEQUENCE 92 AA; 10041 MW; 7BDB64290196CFD2 CRC64;

Query Match 35.7%; Score 5; DB 2; Length 92;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 DGNTL 13
 Db 63 DGNTL 67

RESULT 11
 O62727 PRELIMINARY; PRT; 93 AA.
 AC 062727;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DE ATRIUM L-TYPE CALCIUM CHANNEL (FRAGMENT).
 GN ICA.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OX Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yue L., Wang Z., Nattels S.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF048752; AAC05380.1; -;
 FT NON_TER 1
 FT NON_TER 93
 SQ SEQUENCE 93 AA; 10416 MW; 364B1D46FC8C535 CRC64;

Query Match 35.7%; Score 5; DB 6; Length 93;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 PGWDG 10
 Db 56 PGWDG 60

RESULT 12
 P93074 PRELIMINARY; PRT; 104 AA.
 AC P93074;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PHYTOCHROME (FRAGMENT).
 GN PHYC.
 OS Bambusa sp. (bamboo).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Bambusoideae; Bambuseae; Bambusa.
 OX NCBI_TaxID=48938;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97019052; PubMed-8865668;
 RA Mathews S., Sharrock R.A.;
 RT "The phytochrome gene family in grasses (Poaceae): a phylogeny and
 evidence that grasses have a subset of the loci found in dicot
 angiosperms.";
 RL Mol. Biol. Evol. 13:1141-1150(1996).
 DR EMBL; U61189; AAB41983.1; -;
 DR Mendel; 30961; Bamsp; 2331; 30961.
 DR InterPro; IPR001294; Phytochrome.
 DR InterPro; IPR003018; GAF.
 DR PROSITE; PS00245; PHYTOCHROME_1; 1.
 DR PROSITE; PS00046; PHYTOCHROME_2; 1.
 KW Phytochrome.
 FT NON_TER 1
 FT NON_TER 104
 SQ SEQUENCE 104 AA; 11473 MW; C42A5C80874EA01E CRC64;

Query Match 35.7%; Score 5; DB 10; Length 104;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 DGNTL 13
 Db 75 DGNTL 79

RESULT 13
 Q27100 PRELIMINARY; PRT; 116 AA.
 AC Q27100;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE UBIQUITIN (FRAGMENT).
 GN UB2B.
 OS Trichomonas vaginalis.
 OC Eukaryota; Parabasalidea; Trichomonadida; Trichomonas.
 OX NCBI_TaxID=5722;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NIH-C1, ATCC 30001;
 RX MEDLINE-96081479; PubMed-7490769;
 RA Keeling P.J., Doolittle W.F.;
 RT "Concerted evolution in protists: recent homogenization of a
 polyubiquitin gene in Trichomonas vaginalis.";
 RL J. Mol. Evol. 41:556-562(1995).
 DR EMBL; U28012; AAC46940.1; -;
 DR HSSP; P02248; IUBI
 DR InterPro; IPR000626; Ubiquitin.
 DR Pfam; PF00240; ubiquitin_2.
 DR PRINTS; PR00348; UBIQUITIN.
 DR SMART; SM00213; UBO; 2.
 DR PROSITE; PS00269; UBIQUITIN_1; 1.
 DR PROSITE; PS00053; UBIQUITIN_2; 2.

KW Nuclear protein; Ubiquitin conjugation.

FT NON_TER 1 116
 FT NON_TER 116 116
 SQ SEQUENCE 116 AA; 13155 MW; 2DDF3A7CC4DFF75A CRC64;

Query Match 35.7%; Score 5; DB 5; Length 116;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 DGNL 13

Db 46 DGNL 50

RESULT 14

O9KUE1 PRELIMINARY; PRT; 122 AA.

AC O9KUE1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL PROTEIN VC0580.
 GN VC0580.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae".
 RL Nature 406:477-483(2000).
 DR EMBL; AE004143; AAF93748.1; -.
 DR TIGR; VC0580; -.
 DR InterPro; IPR003509; UPF0102.
 DR Pfam; PF02021; UPF0102; 1.
 KW Complete proteome.
 SQ SEQUENCE 122 AA; 14186 MW; 86F1A87329669BC2 CRC64;

Query Match 35.7%; Score 5; DB 2; Length 122;

Best Local Similarity 100.0%; Pred. No. 83;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 DGNL 13

Db 46 DGNL 50

RESULT 15

O9A8K7 PRELIMINARY; PRT; 131 AA.

AC O9A8K7;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MUTT/NUDIX FAMILY PROTEIN.
 GN CC1346.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005810; AAK23327.1; -.
 DR TIGR; CCI346; -.
 KW Complete proteome.
 SQ SEQUENCE 131 AA; 13994 MW; 5CFDB4354DCF1491 CRC64;

Query Match 35.7%; Score 5; DB 2; Length 131;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AWVDP 6

Db 103 AWVDP 107

Search completed: January 31, 2002, 13:37:40

Job time: 165 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:16 ; Search time 46.78 Seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-73
Perfect score: 14
Sequence: 1 FAWVDPGWDGNTLM 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	11	78.6	154	1 DCD_PYRAB	Q9UXS8 pyrococcus
2	11	78.6	156	1 DCD_PYRHO	O57706 pyrococcus
3	6	42.9	557	1 CO9_RABIT	P48747 myctolagus
4	6	42.9	606	1 ACEA_MYCLE	P46831 mycobacteri
5	5	35.7	35	1 LCGB_LACIA	P36962 lactococcus
6	5	35.7	130	1 RSL1_THEMA	O9X114 thermotoga
7	5	35.7	131	1 PRO2_PARU	Q8T0M8 parietaria
8	5	35.7	141	1 YEDD_SALTY	Q06399 salmonella
9	5	35.7	142	1 RL25_KLULA	P48045 kluyveromyc
10	5	35.7	142	1 RL25_PICJA	P08792 pichia jadi
11	5	35.7	145	1 OAZ_ONCVO	Q9NHZ5 onchocerca
12	5	35.7	171	1 MOUB_ORNMO	O04669 ornithodora
13	5	35.7	171	1 PACA_RANRI	Q91169 r glucagon-
14	5	35.7	175	1 PACA_CHICK	P41534 g glucagon-
15	5	35.7	185	1 SCPI_BRALA	P04569 branchiosto
16	5	35.7	185	1 SCP2_BRALA	P04570 branchiosto
17	5	35.7	199	1 NHAA_PSECL	P27764 pseudomonas
18	5	35.7	206	1 NHAA_RHORH	P29378 rhodococcus
19	5	35.7	209	1 CRBD_BOVIN	P11842 bos taurus
20	5	35.7	210	1 NHAA_PSEPU	P37051 pseudomonas
21	5	35.7	214	1 EOT2_ACTEQ	P17723 actinia equ
22	5	35.7	214	1 EOT4_ACTEQ	Q9Y1U9 actinia equ
23	5	35.7	214	1 EOT5_ACTEQ	Q93109 actinia equ
24	5	35.7	266	1 DHPS_STRPY	O33724 streptococc
25	5	35.7	276	1 KRE9_CANGA	O74683 candida gla
26	5	35.7	289	1 IOLH_BACSU	P42418 bacillus su
27	5	35.7	312	1 GBLP_LEICH	Q27434 leishmania
28	5	35.7	312	1 GBLP_LEIMA	Q25306 leishmania
29	5	35.7	321	1 ASGX_ECOLI	P37595 escherichia
30	5	35.7	326	1 TMOF_PSEME	Q03304 pseudomonas
31	5	35.7	331	1 G3P_MONAN	P53430 monascus an
32	5	35.7	336	1 TS2_NAIZE	P50160 zea mays (m
33	5	35.7	348	1 YE91_SYNY3	P74598 synechocyst

RESULT 1

ID	DCD_PYRAB	STANDARD	PRT	154 AA
AC	O9UXS8			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE)			
GN	DCD OR PAB1164			
OS	Pyrococcus abyssi			
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus			
OX	NCBI_TaxID=25292			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ORSAY			
RA	Heilig R.			
RT	"Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).			
CC	-1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: AJ248288; CAB50685.1; -			
DR	InterPro: IPR003232; dCTP_deaminase			
DR	InterPro: IPR001428; dUTPase			
DR	Pfam: PF00692; dUTPase; 1			
DR	ProDom: PD004900; dCTP_deaminase; 1			
KW	Hydrolase; Complete proteome.			
SQ	SEQUENCE 154 AA; 17758 MW; C031BEE419094DDB CRC64;			

Query Match 78.6% Score 11; DB 1; Length 154;

Best Local Similarity 100.0%; Pred. No. 7.1e-06; Mismatches 0; Indels 0; Gaps 0;

Matches 11; Conservative 0;

Qy 1 FAWVDPGWDGN 11
|||||

Db 91 FAWVDPGWDGN 101
|||||

RESULT 2

ID	DCD_PYRHO	STANDARD	PRT	156 AA
AC	O57706			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			

DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR PH1997.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka K., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-75(1998).
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O -> DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP000007; BAA31124.1;
DR InterPro: IPR003232; dCTP_deaminase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 156 AA; 17871 MW; 96B2C2C50393A985 CRC64;

Query Match 78.6%; Score 11; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 7.2e-06;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FAWVDPGWGDN 11
DB 91 FAWVDPGWGDN 101
|||||
[1]

RESULT 3
ID CO9_RABIT STANDARD; PRT; 557 AA.
AC P48747;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COMPLEMENT COMPONENT C9 PRECURSOR.
GN C9.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-34.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX MEDLINE=95181293; PubMed=7533152;
RA Huesler T., Lockert D.H., Kaufman K.M., Soderz J.M., Sims P.J.;
RT "Chimeras of human complement C9 reveal the site recognized by
complement regulatory protein CD59.";
RL J. Biol. Chem. 270:3483-3486(1995).
CC -1- FUNCTION: C9 IS THE FINAL COMPONENT OF THE COMPLEMENT SYSTEM TO
CC BE ADDED IN THE ASSEMBLY OF THE MEMBRANE ATTACK COMPLEX. IT IS
CC ABLE TO ENTER LIPID BILAYERS, FORMING TRANSMEMBRANE CHANNELS.
CC
CC -1- PTM: THROMBIN CLEAVES FACTOR C9 TO PRODUCE C9A AND C9B.

CC -1- SIMILARITY: TO COMPLEMENT FACTORS C6, C7, C8, AND TO PERFORIN.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U20055; AAC48459.1;
DR HSP: P01130; IAJJ
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002172; LDL_recept-A.
DR InterPro: IPR001862; MAC_perforin.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00057; LDL_recept_a; 1.
DR Pfam: PF01823; MACPF; 1.
DR Pfam: PF00080; Tsp_1; 1.
DR PRINTS: PR00764; COMPLEMENTC9.
DR SMART: SM00192; LDLa; 1.
DR SMART: SM00457; MACPF; 1.
DR SMART: SM00209; TSPI; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01209; LDLRA_1; 1.
DR PROSITE: PS00068; LDLRA_2; 1.
DR PROSITE: PS00279; MAC_PERFORIN; 1.
DR PROSITE: PS00092; TSPI; 1.
KW Complement pathway; Complement alternate pathway; Glycoprotein;
KW Plasma; Membrane attack complex; Cytolysis; Transmembrane; Signal;
KW EGF-like domain.
FT CHAIN 1 21
FT DOMAIN 42 98
FT TRANSMEM 319 335
FT TRANSMEM 340 359
FT DOMAIN 516 550
FT DISULFID 101 113
FT DISULFID 108 126
FT DISULFID 120 135
FT DISULFID 385 414
FT DISULFID 520 536
FT DISULFID 523 538
FT DISULFID 540 549
FT CARBOHYD 261 261
FT CARBOHYD 282 282
FT CARBOHYD 424 424
SQ SEQUENCE 557 AA; 62662 MW; FF4E65FF8D1AB417 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 557;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 WDGNLT 13
DB 186 WDGNLT 191
|||||
[1]

RESULT 4
ID ACEA_MYCLE STANDARD; PRT; 606 AA.
AC P46831; O9CBH0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ISOCITRATE LYASE (EC 4.1.3.1) (ISOCITRASE) (ISOCITRASE) (ICL).
GN ACEA OR ML1985.

OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96059637; PubMed=7476188;
RA Fsihi H., Cole S.T.;
RT "The Mycobacterium leprae genome: systematic sequence analysis
RT identifies key catabolic enzymes, ATP-dependent transport systems and
RT a novel pola locus associated with genomic variability.";
RL Mol. Microbiol. 16:909-919(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RA MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parthill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier C., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: ISOCITRATE = SUCCINATE + GLYOXYLATE.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS (BY SIMILARITY).
CC -!- PATHWAY: FIRST STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE
CC TRICARBOXYLIC ACID CYCLE (IN BACTERIA, PLANTS, AND FUNGI).
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE ISOCITRATE LYASE FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 569.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z46257; CAA86357.1;
DR EMBL; AL583924; CAC30940.1; ALT_FRAME.
DR Leproma; ML1985;...
DR InterPro; IPR000918; Isocit_lyase.
DR Pfam; PF00463; ICL; 2.
DR ProDom; PD001857; Isocit_lyase; 1.
DR PROSITE; PS00161; ISOCITRATE_LYASE; 1.
KW Glyoxylate bypass; Tricarboxylic acid cycle; Lyase; Phosphorylation;
KW Complete proteome. 211
FT ACT_SITE 211 BY SIMILARITY.
FT CONFLICT 158 R -> L (IN REF. 1).
FT CONFLICT 569 V -> L (IN REF. 1).
SQ SEQUENCE 606 AA; 67600 MW; D82CCADD6B6D384 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VDPGWD 9
Db 413 VDPGWD 418
|||||

RESULT 5
LCGB_LACLA
ID LCGB_LACLA STANDARD; PRT; 35 AA.
AC P36962;

Mycobacterium leprae.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1769;
[1]
SEQUENCE FROM N.A.
MEDLINE=96059637; PubMed=7476188;
Fsihi H., Cole S.T.;
"The Mycobacterium leprae genome: systematic sequence analysis
identifies key catabolic enzymes, ATP-dependent transport systems and
a novel pola locus associated with genomic variability.";
Mol. Microbiol. 16:909-919(1995).
[2]
SEQUENCE FROM N.A.
STRAIN=TN;
MEDLINE=21128732; PubMed=11234002;
Cole S.T., Eigmeier K., Parthill J., James K.D., Thomson N.R.,
Wheeler P.R., Honore N., Garnier C., Churcher C., Harris D.,
Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,
Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
Barrell B.G.;
"Massive gene decay in the leprosy bacillus.";
Nature 409:1007-1011(2001).
-!- CATALYTIC ACTIVITY: ISOCITRATE = SUCCINATE + GLYOXYLATE.
-!- COFACTOR: REQUIRES DIVALENT CATIONS (BY SIMILARITY).
-!- PATHWAY: FIRST STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE
TRICARBOXYLIC ACID CYCLE (IN BACTERIA, PLANTS, AND FUNGI).
-!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: CYTOPLASMIC.
-!- SIMILARITY: BELONGS TO THE ISOCITRATE LYASE FAMILY.
-!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 569.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; Z46257; CAA86357.1;
EMBL; AL583924; CAC30940.1; ALT_FRAME.
Leproma; ML1985;...
InterPro; IPR000918; Isocit_lyase.
Pfam; PF00463; ICL; 2.
ProDom; PD001857; Isocit_lyase; 1.
PROSITE; PS00161; ISOCITRATE_LYASE; 1.
Glyoxylate bypass; Tricarboxylic acid cycle; Lyase; Phosphorylation;
Complete proteome. 211
ACT_SITE 211 BY SIMILARITY.
CONFLICT 158 R -> L (IN REF. 1).
CONFLICT 569 V -> L (IN REF. 1).
SEQUENCE 606 AA; 67600 MW; D82CCADD6B6D384 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 VDPGWD 9
Db 413 VDPGWD 418
|||||

RESULT 5
LCGB_LACLA
ID LCGB_LACLA STANDARD; PRT; 35 AA.
AC P36962;

DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE BACTERIOCIN LACTOCOCIN G, BETA SUBUNIT
OC Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE.
RX STRAIN=LMG 2081;
RA MEDLINE=92380946; PubMed=1512201;
RA Nissen-Meyer J., Holo H., Havarstein L.S., Sletten K., Nes I.F.;
RT "A novel lactococcal bacteriocin whose activity depends on the
RT complementary action of two peptides.";
RL J. Bacteriol. 174:5686-5692(1992).
CC -!- FUNCTION: KILLS LACTOCOCCI.
CC -!- SUBUNIT: BACTERIOCIN ACTIVITY REQUIRES INTERACTION OF ALPHA AND
CC BETA PEPTIDES IN A MOLAR RATIO OF 7:1 OR 8:1 RESPECTIVELY.
DR PIR; C44918; C44918.
KW Antibiotic; Bacteriocin.
SQ SEQUENCE 35 AA; 4110 MW; 76109F8BB0C489D2 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AWVDP 6
Db 7 AWVDP 11
|||||

RESULT 6
RS11_THEME
ID RS11_THEME STANDARD; PRT; 130 AA.
AC Q9X1I4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S11.
GN RPSK OR TM1474.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MSB8 / DSM 3109;
RA MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Raft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Raizberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- FUNCTION: S11 PLAYS AN ESSENTIAL ROLE FOR THE SELECTION OF THE
CC CORRECT TRNA IN PROTEIN BIOSYNTHESIS. IT IS LOCATED ON THE LARGE
CC LOBE OF THE SMALL SUBUNIT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S11P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AE001798; AAD36542.1;
DR TIGR; TM1474;

```
DR InterPro; IPR001971; Ribosomal_S11.
DR Pfam; PF00411; Ribosomal_S11; 1.
DR PRODOM; PD001010; Ribosomal_S11; 1.
DR PROSITE; PS00054; RIBOSOMAL_S11; FALSE_NEG.
SQ Ribosomal protein; Complete proteome.
KW RIBOSOMAL_S11; 14092 MW; E2D826640FE037CA CRC64;
SEQUENCE 130 AA; 14092 MW; E2D826640FE037CA CRC64;

Query Match 35.7%; Score 5; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
Db 39 DGNTL 43

RESULT 7
PROZ_PARJU STANDARD; PRT; 131 AA.
AC Q9T0M8; 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
GN PROFILIN 2 (POLLEN ALLERGEN PAR J 3).
OS Parietaria judaica.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids I; Rosales; Urticaceae; Parietaria.
OX NCBI_TaxID=33127;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pollen;
RA Asturias J.A., Arilla M.C., Gomez-Bayon N., Martinez A., Martinez J.,
RA Palacios R.;
RT "Analysis of recombinant allergen Par j 3 (profilin) from Parietaria
RT judaica";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS TO ACTIN AND AFFECTS THE STRUCTURE OF THE
CC CYTOSKELETON. AT HIGH CONCENTRATIONS, PROFILIN PREVENTS THE
CC POLYMERIZATION OF ACTIN, WHEREAS IT ENHANCES IT AT LOW
CC CONCENTRATIONS. BY BINDING TO PIP2, IT INHIBITS THE FORMATION OF
CC IP3 AND DG (BY SIMILARITY).
CC -!- SUBUNIT: OCCURS IN MANY KINDS OF CELLS AS A COMPLEX WITH MONOMERIC
CC ACTIN IN A 1:1 RATIO.
CC -!- DISEASE: CAUSE ALLERGIC REACTIONS.
CC -!- SIMILARITY: BELONGS TO THE PROFILIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y15209; CAB44257.1; -.
CC HSSP; P25816; IQQA.
CC InterPro; IPR002097; Profilin.
CC Pfam; PF00235; profilin; 1.
CC PRINTS; PR00392; PROFILIN.
CC SMART; SM00392; PROF; 1.
CC PROSITE; PS00414; PROFILIN; 1.
KW Actin-binding; Cytoskeleton; Multigene family; Allergen.
SQ SEQUENCE 131 AA; 13784 MW; 8A4FC0BAC1565AB CRC64;

Query Match 35.7%; Score 5; DB 1; Length 131;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
Db 39 DGNTL 43

RESULT 8
YEDD_SALTY STANDARD; PRT; 141 AA.
AC Q06399; 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
GN HYPOTHETICAL PROTEIN YEDD.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JW1103;
RA Raha M., Kihara M., Kawagishi I., Macnab R.M.;
RT "Organization of the Escherichia coli and Salmonella typhimurium
RT chromosomes between flagellar regions IIIa and IIb, including a
RT large non-coding region.";
RL J. Gen. Microbiol. 139:1401-1407(1993).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L13280; AAA71971.1; -.
CC StyGene; SG10420; yedd.
KW Hypothetical protein.
SQ SEQUENCE 141 AA; 15465 MW; A37E9E11ADC8BD9 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DGNTL 13
Db 102 DGNTL 106

RESULT 9
RL25_KLULA STANDARD; PRT; 142 AA.
AC P48045; 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
GN 60S RIBOSOMAL PROTEIN L25.
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=9312726; PubMed=1481569;
RA Bergkamp-Steffens G.K., Hoekstra R., Planta R.J.;
RT "Structural and putative regulatory sequences of Kluyveromyces
RT ribosomal protein genes.";
RL Yeast 8:903-922(1992).
CC -!- SIMILARITY: BELONGS TO THE L23P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; S53176; AAB24896.1; -
DR InterPro; IPR001014; Ribosomal_L23.
DR Pfam; PF00276; Ribosomal_L23; 1.
DR ProDom; PD001141; Ribosomal_L23; 1.
DR PROSITE; PS00050; RIBOSOMAL_L23; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 142 AA; 15716 MW; E8F8EE77D344009C CRC64;

Query Match 35.7%; Score 5; DB 1; Length 142;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTL 13
Db 78 DGNTL 82
|||||

RESULT 10
RL25_PICJA STANDARD; PRT; 142 AA.
AC P08792;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L25.
GN RPL25.
OS Pichia jadinii (Yeast) (Candida utilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4903;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88210534; PubMed-3449224;
RA Woudt L.P., Mager W.H., Beek J.G., Wassenaar G.M., Planta R.J.;
RT "Structural and putative regulatory sequences of the gene encoding
RT ribosomal protein L25 in Candida utilis.";
RL Curr. Genet. 12:193-198(1987).
CC -1- FUNCTION: THIS PROTEIN BINDS TO A SPECIFIC REGION ON THE 26S RRNA.
CC -1- SIMILARITY: BELONGS TO THE L23P FAMILY OF RIBOSOMAL PROTEINS.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X05919; CAA29354.1; -
DR PIR; S06373; R5HQ25.
DR InterPro; IPR001014; Ribosomal_L23.
DR Pfam; PF00276; Ribosomal_L23; 1.
DR ProDom; PD001141; Ribosomal_L23; 1.
DR PROSITE; PS00050; RIBOSOMAL_L23; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 142 AA; 15632 MW; 6FDA79C48A9B06B3 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 142;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTL 13
Db 78 DGNTL 82
|||||

RESULT 11
OAZ_ONCVO STANDARD; PRT; 145 AA.
AC Q9NH25;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ORNITHINE DECARBOXYLASE ANTIZYME (ODC-AZ).
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20237626; PubMed=10775274;
RA Ivanov I.P., Matsufuji S., Murakami Y., Gesteland R.F., Atkins J.F.;
RT "Conservation of polyamine regulation by translational frameshifting
RT from yeast to mammals.";
RL EMBO J. 19:1907-1917(2000).

CC -1- FUNCTION: BINDS TO, AND DESTABILIZES, ORNITHINE DECARBOXYLASE.
CC -1- MISCELLANEOUS: A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS
CC FOR SER-28 AND ASP-29. AN AUTOREGULATORY MECHANISM ENABLES
CC MODULATION OF FRAMESHIFTING ACCORDING TO THE CELLULAR
CC CONCENTRATION OF POLYAMINES.
CC -1- SIMILARITY: BELONGS TO THE ODC ANTIZYME FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF217279; AAF68270.1; -
DR InterPro; IPR002993; ODC_AZ.
DR Pfam; PF02100; ODC_AZ; 1.
DR PROSITE; PS01337; ODC_AZ; 1.
KW Ribosomal frameshift.
SQ SEQUENCE 145 AA; 16199 MW; 08F1FB94766BEF63 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 DGNTL 13
Db 61 DGNTL 65
|||||

RESULT 12
MOUB_ORNMO STANDARD; PRT; 171 AA.
AC O04659;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MOUBATIN PRECURSOR.
OS Ornithodoros moubata (Soft tick).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Parasitiformes; Ixodida; Argasidae; Ornithodoros.
OX NCBI_TaxID=6938;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE-93194835; PubMed=8449907;
RA Keller P.M., Waxman L., Arnold B.A., Schultz L.D., Condra C.,
RA Connolly T.M.;
RT "Cloning of the cDNA and expression of moubatin, an inhibitor of
RT platelet aggregation.";
RL J. Biol. Chem. 268:5450-5456(1993).

Regul. Pept. 37:326-326(1992).
 CC -1- FUNCTION: PRIMARY ROLE OF GRF IS TO RELEASE GH FROM THE
 CC PITUITARY.
 CC -1- FUNCTION: PACAP PLAYS PIVOTAL ROLES AS A NEUROTRANSMITTER AND/OR
 CC A NEUROMODULATOR.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GRF 1-43, GRF 1-46 (SHOWN
 CC HERE) AND GRF 33-46; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U71183; AAB51200.1; -;
 DR EMBL; U71184; AAB51201.1; -;
 DR EMBL; U71185; AAB51202.1; -;
 DR PIR; A61070; A61070.
 DR Pfam; PF00123; hormone2; 2.
 DR PRINTS; PR00275; GLUCAGON.
 DR SMART; SM00070; GLUCA; 2.
 DR PROSITE; PS00260; GLUCAGON; 2.
 DR Glucagon family; Hormone; Cleavage on pair of basic residues; Signal;
 KW Amidation; Alternative splicing.
 FT SIGNAL 1 23 POTENTIAL.
 FT PROPEP 24 80
 FT PEPTIDE 83 128
 FT PEPTIDE 131 168
 FT PEPTIDE 131 157
 FT MOD_RES 157 157
 FT MOD_RES 168 168
 FT VARSPLIC 82 114
 FT VARSPLIC 115 117
 FT VARSPLIC 175 AA; 19560 MW; 0DB54995F0AA9DFB CRC64;
 SQ SEQUENCE 175 AA; 19560 MW; 0DB54995F0AA9DFB CRC64;
 Query Match 35.7%; Score 5; DB 1; Length 175;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 DGNTL 13
 Db 46 DGNTL 50
 RESULT 15
 SCPL_BRLA STANDARD; PRT; 185 AA.
 AC P04569;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE SARCOPLASMIC CALCIUM-BINDING PROTEINS I, III, AND IV (SCP I, III, IV).
 OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7740;
 RN [1]
 RP SEQUENCE (SCP I).
 RA Takagi T., Konishi K., Cox J.A.;
 RT "Amino acid sequence of two sarcoplasmic calcium-binding proteins
 RT from the protochordate amphioxus.";
 RL Biochemistry 25:3585-3592(1986).
 RN [2]
 RP SEQUENCE (SCP III AND IV), AND REVISIONS (SCP I).
 RX MEDLINE=91006125; PubMed=2209593;

RA Takagi T., Cox J.A.;
 RT "Amino acid sequences of four isoforms of amphioxus sarcoplasmic
 RT calcium-binding proteins.";
 RL Eur. J. Biochem. 192:387-399(1990).
 CC -1- FUNCTION: LIKE PARVALBUMINS, SCP'S SEEM TO BE MORE ABUNDANT IN
 CC FAST CONTRACTING MUSCLES, BUT NO FUNCTIONAL RELATIONSHIP CAN BE
 CC ESTABLISHED FROM THIS DISTRIBUTION.
 CC -1- MISCELLANEOUS: THE SARCOPLASMIC CALCIUM-BINDING PROTEINS ARE
 CC ABUNDANT IN THE MUSCLE OF ARTHROPODS, MOLLUSCS, ANNELIDS, AND
 CC PROTOCHORDATES.
 CC -1- MISCELLANEOUS: THIS PROTEIN HAS THREE FUNCTIONAL CALCIUM-BINDING
 CC SITES: POTENTIAL SITE 4 HAS LOST AFFINITY FOR CALCIUM.
 CC -1- MISCELLANEOUS: THERE ARE 7 DIFFERENT SCP'S IN AMPHIOXUS. THE
 CC SEQUENCE SHOWN HERE IS THAT OF SCP I.
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
 DR PIR; B24479; B24479.
 DR PIR; S13182; S13182.
 DR PIR; S13184; S13184.
 DR PIR; S13185; S13185.
 DR HSP; P04570; 2SAS.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; efhand; 3.
 DR PROSITE; PS00018; EF_HAND; 3.
 KW Muscle protein; Calcium-binding; Duplication.
 FT CA_BIND 19 30 EF-HAND 1.
 FT CA_BIND 70 81 EF-HAND 2.
 FT CA_BIND 115 126 EF-HAND 3.
 FT DOMAIN 150 161 ANCESTRAL CALCIUM SITE 4.
 FT VARIANT 20 20 Y -> M (IN SCP III).
 FT VARIANT 23 23 D -> N (IN SCP IV).
 SQ SEQUENCE 185 AA; 21418 MW; 92D556A663003FAD CRC64;
 Query Match 35.7%; Score 5; DB 1; Length 185;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 GNTLM 14
 Db 177 GNTLM 181

Search completed: January 31, 2002, 13:39:18
 Job time: 78 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:09 ; Search time 65.13 seconds
(without alignments)
4.492 Million cell updates/sec

Title: US-08-957-709-72
Perfect score: 13
Sequence: 1 XGXXDXGXXXXX 13

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCRU5_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2	15.4	11	4	US-09-021-247-16
2	2	15.4	12	1	US-07-616-910-29
3	2	15.4	12	5	PCT-US91-08497-29
4	2	15.4	14	4	US-08-822-774-53
5	2	15.4	19	3	US-08-802-981-52
6	2	15.4	23	4	US-08-669-304-2
7	2	15.4	23	4	US-08-669-304-9
8	2	15.4	40	5	PCT-US95-10973A-77
9	2	15.4	68	2	US-08-511-485-2
10	2	15.4	68	3	US-09-212-971-2
11	2	15.4	68	4	US-08-800-929A-2
12	2	15.4	68	4	US-09-617-053A-2
13	2	15.4	273	4	US-08-928-213B-10
14	2	15.4	455	1	US-07-762-681A-2
15	2	15.4	526	1	US-08-895-590-5
16	1	7.7	2	1	US-07-820-154A-15
17	1	7.7	2	1	US-07-820-154A-23
18	1	7.7	2	1	US-07-820-154A-33
19	1	7.7	2	1	US-07-820-154A-39
20	1	7.7	2	1	US-07-791-213D-39
21	1	7.7	2	1	US-07-729-353-5
22	1	7.7	2	1	US-08-122-510-9
23	1	7.7	2	1	US-08-122-510-14
24	1	7.7	2	1	US-08-358-160-171
25	1	7.7	2	1	US-08-153-799-20
26	1	7.7	2	1	US-08-191-866D-69
27	1	7.7	2	1	US-08-293-150A-39

28	1	7.7	2	1	US-08-470-837-6	Sequence 6, Appli
29	1	7.7	2	2	US-08-272-255-2	Sequence 2, Appli
30	1	7.7	2	2	US-08-097-554A-15	Sequence 15, Appl
31	1	7.7	2	2	US-08-097-554A-23	Sequence 23, Appl
32	1	7.7	2	2	US-08-097-554A-33	Sequence 33, Appl
33	1	7.7	2	2	US-08-097-554A-39	Sequence 39, Appl
34	1	7.7	2	2	US-08-539-432-15	Sequence 15, Appl
35	1	7.7	2	2	US-08-539-432-16	Sequence 16, Appl
36	1	7.7	2	2	US-08-539-432-17	Sequence 17, Appl
37	1	7.7	2	2	US-08-185-949B-69	Sequence 69, Appl
38	1	7.7	2	2	US-08-483-236-4	Sequence 4, Appli
39	1	7.7	2	2	US-09-060-455-1	Sequence 1, Appli
40	1	7.7	2	3	US-09-327-424-2	Sequence 2, Appli
41	1	7.7	2	3	US-08-480-640A-15	Sequence 15, Appl
42	1	7.7	2	3	US-08-480-640A-23	Sequence 23, Appl
43	1	7.7	2	3	US-08-480-640A-33	Sequence 33, Appl
44	1	7.7	2	3	US-08-480-640A-39	Sequence 39, Appl
45	1	7.7	2	3	US-09-039-308A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-021-247-16
; Sequence 16, Application US/09021247
; Patent No. 6225444
; GENERAL INFORMATION:
; APPLICANT: Shashoua, Victor E.
; TITLE OF INVENTION: NEUROPROTECTIVE PEPTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,247
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: N0260/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: Region
; LOCATION: 2
; OTHER INFORMATION: /note= "Xaa = Asp, Asn, Thr or Glu"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 3
; OTHER INFORMATION: /note= "Xaa = any amino acid"
US-09-021-247-16

Query Match 15.4%; Score 2; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GXXD 5
Db 1 GXXD 4

RESULT 2
US-07-616-910-29
; Sequence 29, Application US/07616910
; Patent No. 5223484
; GENERAL INFORMATION:
; APPLICANT: Fleming, Patrick J.
; APPLICANT: Kent, Ute M.
; TITLE OF INVENTION: Peptide Which Regulates Weight Gain in
; MAMMALS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 19901121
; CLASSIFICATION: 514
; APPLICATION NUMBER: US/07/616,910
; FILING DATE: 21-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, Norman F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-069-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US91-08497-29

Query Match 15.4%; Score 2; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GXXD 5
Db 7 GXXD 10

RESULT 3
PCT-US91-08497-29
; Sequence 29, Application PC/TUS9108497
; GENERAL INFORMATION:
; APPLICANT: Fleming, Patrick J.
; APPLICANT: Kent, Ute M.
; TITLE OF INVENTION: Peptide Which Regulates Weight Gain in
; MAMMALS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 19901121
; CLASSIFICATION: 514
; APPLICATION NUMBER: US/07/616,910
; FILING DATE: 21-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, Norman F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 1126-069-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)521-4500
; TELEFAX: (703)486-2347
; TELEX: 248855 OPAT
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US91-08497-29

STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/08497
FILING DATE: 19911121
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/616,910
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, Norman F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-069-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US91-08497-29

Query Match 15.4%; Score 2; DB 5; Length 12;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GXXD 5
Db 7 GXXD 10

RESULT 4
US-08-822-774-53
; Sequence 53, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C. Suite 700
; STREET: 1200 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-822-774-53

Query Match 15.4%; Score 2; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
| | | |
Db 2 GXXD 5

RESULT 5
US-08-802-981-52
; Sequence 52, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: COMPOSITIONS for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-0003000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /product= "A1b"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /product= "Acp"
FEATURE:
NAME/KEY: Modified-site

LOCATION: 5
OTHER INFORMATION: /product= "Acp"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 14
OTHER INFORMATION: /product= "Acp"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 15
OTHER INFORMATION: /product= "Acp"
US-08-802-981-52

Query Match 15.4%; Score 2; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DXXGX 9
| | | | |
Db 2 DXXGX 6

RESULT 6
US-08-669-304-2
; Sequence 2, Application US/08669304
; Patent No. 6251626
; GENERAL INFORMATION:
; APPLICANT: Peter Stougaard
; APPLICANT: Ole Cal Hansen
; TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
; TITLE OF INVENTION: METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hunton & Williams
; STREET: 1900 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006-1109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,304
; FILING DATE: 12 July 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,910
; FILING DATE: 7 June 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stanislaus Aksman
; REGISTRATION NUMBER: 28,562
; REFERENCE/DOCKET NUMBER: 320.000003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 955-1926
; TELEFAX: (202) 778-2201
; TELEX: NO. 6251626e
; INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-669-304-2

Query Match 15.4%; Score 2; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DXXG 9
Db 14 DXXG 18

RESULT 7
PCT-US95-10973A-9
; Sequence 9, Application US/08669304
; Patent No. 6251626
; GENERAL INFORMATION:
; APPLICANT: Peter Stougaard
; APPLICANT: Ole Cai Hansen
; TITLE OF INVENTION: RECOMBINANT HEXOSE OXIDASE, A
; METHOD OF PRODUCING SAME AND USE OF SUCH ENZYME
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hunton & Williams
; STREET: 1900 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006-1109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,304
; FILING DATE: 12 July 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,910
; FILING DATE: 7 June 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Stanislaus Aksman
; REGISTRATION NUMBER: 26,562
; REFERENCE/DOCKET NUMBER: 320.000003
; TELEPHONE: (202) 935-1926
; TELEFAX: (202) 778-2201
; TELEX: No. 6251626
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-669-304-9

Query Match 15.4%; Score 2; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DXXG 9
Db 14 DXXG 18

RESULT 8
PCT-US95-10973A-77
; Sequence 77, Application PC/TUS9510973A
; GENERAL INFORMATION:
; APPLICANT: Prism Pharmaceuticals, Inc.
; TITLE OF INVENTION: CONJUGATES OF VASCULAR ENDOTHELIAL GROWTH FACTOR WITH TARGET
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10973A
FILING DATE: 29-AUG-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Nottenburg, Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.413PC
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: /product= "SO-4"
PCT-US95-10973A-77

Query Match 15.4%; Score 2; DB 5; Length 40;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GXGD 5
Db 35 GXGD 38

RESULT 9
US-08-511-485-2
; Sequence 2, Application US/08511485
; Patent No. 5919912
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G.
; APPLICANT: Mackenzie, Alexander E.
; APPLICANT: Baird, Stephen
; TITLE OF INVENTION: MAMMALIAN IAP GENE FAMILY, PRIMERS,
; AND DETECTION METHODS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/511.485
FILING DATE: 04-AUG-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 07540/002001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: both
; MOLECULE TYPE: protein
; FEATURE:
; OTHER INFORMATION: Xaa at positions 1, 2, 3, 6, 9, 10, 14, 15, 18, 19, 20, 21, 24, 30, 32, 33, 35, 37, 40,
; OTHER INFORMATION: 42, 43, 44, 45, 46, 47, 49, 50, 51, 53, 54, 55, 56, 57, 59, 6
; OTHER INFORMATION: 61, 62, 64 and 66 may be any amino acid. Xaa at positions 13
; OTHER INFORMATION: be any amino acid or may be absent.
; US-08-511-485-2

Query Match 15.4%; Score 2; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GXXD 5
Db 31 GXXD 34

RESULT 10

US-09-212-971-2
; Sequence 2, Application US/09212971B
; Patent No. 6107041
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
; TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
; TITLE OF INVENTION: DISEASE
; FILE REFERENCE: 07891/009002
; CURRENT APPLICATION NUMBER: US/09/212,971B
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 60/017,354
; EARLIER FILING DATE: 1996-04-26
; EARLIER APPLICATION NUMBER: 60/030,590
; EARLIER FILING DATE: 1996-11-14
; EARLIER APPLICATION NUMBER: 08/800,929
; EARLIER FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Xaa at 1, 2, 3, 6, 9, 10, 14, 15, 18, 19, 20, 21,
; OTHER INFORMATION: 24, 30, 32, 33, 35, 37, 40, 42, 43, 44, 45, 46,
; OTHER INFORMATION: 47, 49, 50, 51, 53, 54, 55, 56, 57, 59, 60, 61,
; OTHER INFORMATION: 62, 64 and 66 can be any amino acid; Xaa at 13, 16
; OTHER INFORMATION: and 17 can be any amino acid or absent.
; OTHER INFORMATION: Based on consensus from Homo sapiens and Mus
; OTHER INFORMATION: musculus
; US-09-212-971-2

Query Match 15.4%; Score 2; DB 3; Length 68;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GXXD 5
Db 31 GXXD 34

RESULT 11

US-08-800-929A-2
; Sequence 2, Application US/08800929A
; Patent No. 6133437
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF
; TITLE OF INVENTION: IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERAT
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,929A
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/030,590
; FILING DATE: 14-NOV-1996
; APPLICATION NUMBER: 60/017,354
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bieker-Brady, Kristina
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 07891/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Other
; LOCATION: 1...68
; OTHER INFORMATION: Xaa can be any amino acid.
; US-08-800-929A-2

Query Match 15.4%; Score 2; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GXXD 5
Db 31 GXXD 34

RESULT 12

US-09-617-053A-2
; Sequence 2, Application US/09617053A
; Patent No. 6300492

GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G
APPLICANT: Mackenzie, Alexander E
APPLICANT: Liston, Peter
APPLICANT: Baird, Stephen
APPLICANT: Tsang, Benjamin K
APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPs AND
TITLE OF INVENTION: NACS FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
FILE REFERENCE: 07891/009003
CURRENT APPLICATION NUMBER: US/09/617,053A
CURRENT FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/800,929
PRIOR FILING DATE: 1997-02-13
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 68
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Xaa at 1, 2, 3, 6, 9, 10, 14, 15, 18, 19, 20, 21,
OTHER INFORMATION: 24, 30, 32, 33, 35, 37, 40, 42, 43, 44, 45, 46,
OTHER INFORMATION: 47, 49, 50, 51, 53, 54, 55, 56, 57, 59, 60, 61,
OTHER INFORMATION: 62, 64 and 66 can be any amino acid; Xaa at 13, 16
OTHER INFORMATION: and 17 can be any amino acid or absent.
OTHER INFORMATION: Based on consensus from Homo sapiens and Mus
OTHER INFORMATION: musculus
US-09-617-053A-2

Query Match 15.4%; Score 2; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXD 5
DB 31 GXD 34

RESULT 13
US-08-928-213B-10
Sequence 10, Application US/08928213B
Patent No. 6238905
GENERAL INFORMATION:
APPLICANT: McHenry, Charles S.
Seville, Mark
Cull, Millard G.
TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III
HOLOEZYME
NUMBER OF SEQUENCES: 195
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,213B
FILING DATE: 12-Sep-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: ENZYCO-02550
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-705-8410
TELEFAX: 415-397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-08-928-213B-10

Query Match 15.4%; Score 2; DB 4; Length 273;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXD 5
DB 242 GXD 245

RESULT 14
US-07-762-681A-2
Sequence 2, Application US/07762681A
Patent No. 5266475
GENERAL INFORMATION:
APPLICANT: Lee, Chanyong
APPLICANT: Bagdasarian, Michael
APPLICANT: Zeikus, J. Gregory
APPLICANT: Meng, Menghsiao
TITLE OF INVENTION: Glucose Isomerases With Improved
TITLE OF INVENTION: Affinity For D-Glucose
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thad Kryshak
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/762,681A
FILING DATE: 19910919
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kryshak, Thad F
REGISTRATION NUMBER: 19,428
REFERENCE/DOCKET NUMBER: 66-336-9038-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 455 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-07-762-681A-2

Query Match 15.4%; Score 2; DB 1; Length 455;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DXXXG 9

Db 352 DXXG 356
|||||

RESULT 15
US-08-895-590-5
; Sequence 5, Application US/08895590
; Patent No. 6207410
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,590
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,888
; FILING DATE: 19-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-263
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 526 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-895-590-5

Query Match 15.4%; Score 2; DB 4; Length 526;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
|||||
Db 75 GXXD 78

Search completed: January 31, 2002, 13:15:10
Job time: 91 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:05 ; Search time 140.03 Seconds
(without alignments)
6.877 Million cell updates/sec

Title: US-08-957-709-72
Perfect score: 13
Sequence: 1 XGXXDXXXXXXX 13

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- A_Geneseq.ll01.*
- 1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2	15.4	4	22	AA74959
2	2	15.4	9	16	AA75127
3	2	15.4	11	16	AA86325
4	2	15.4	11	20	AA745210
5	2	15.4	12	13	AA24843
6	2	15.4	14	19	AA72848
7	2	15.4	14	22	AA62873
8	2	15.4	15	20	AA42757
9	2	15.4	15	20	AA39335
10	2	15.4	19	19	AA82117
11	2	15.4	20	19	AA82125

12	2	15.4	20	22	AA76715	Human colon cancer
13	2	15.4	20	22	AA73106	Protease binding s
14	2	15.4	20	22	AA81828	Human endoepine-l
15	2	15.4	20	22	AA81831	Human endoepine-l
16	2	15.4	23	18	AA20069	HOX2, a hexose oxi
17	2	15.4	24	21	AA29933	Human CTLA-4 scaff
18	2	15.4	39	21	AA15187	Arabidopsis thalia
19	2	15.4	40	9	AA81583	N-terminal of ribo
20	2	15.4	42	22	AA74971	Human colon cancer
21	2	15.4	52	21	AA15186	Arabidopsis thalia
22	2	15.4	58	21	AA60442	Human secreted pro
23	2	15.4	59	21	AA165156	Human 5' EST relat
24	2	15.4	68	18	AA19588	Mammalian inhibito
25	2	15.4	73	21	AA24895	Arabidopsis thalia
26	2	15.4	74	21	AA61652	zea mays protein f
27	2	15.4	76	16	AA84516	Hepatitis C virus
28	2	15.4	86	7	AA60952	Sequence of endoge
29	2	15.4	104	16	AA84352	Hepatitis C virus
30	2	15.4	111	20	AA19730	SEQ ID NO 448 from
31	2	15.4	119	21	AA64868	zea mays protein f
32	2	15.4	123	22	AA72675	Murine OR-like pol
33	2	15.4	130	22	AA72677	Human colon cancer
34	2	15.4	150	20	AA27137	Vertebrate slit pr
35	2	15.4	160	21	AA33833	Arabidopsis thalia
36	2	15.4	161	22	AA33834	Human gene 7 encod
37	2	15.4	166	21	AA33832	Arabidopsis thalia
38	2	15.4	168	22	AA63637	Human gastric canc
39	2	15.4	178	22	AA75792	Human colon cancer
40	2	15.4	194	22	AA603407	Human gene 7 encod
41	2	15.4	218	21	AA58877	Breast and ovarian
42	2	15.4	455	14	AA44236	Glucose isomerase
43	2	15.4	626	20	AA705852	Banana ripening fr
44	2	15.4	797	20	AA705849	Banana ripening fr
45	2	15.4	1626	22	AA72447	UGT consensus seq

ALIGNMENTS

RESULT 1	
AA74959	
ID	AA74959 standard; peptide; 4 AA.
XX	
AC	AA74959;
XX	
DT	10-JUL-2001 (first entry)
XX	
DE	Octopus minor brain cardiotoxic generic peptide.
XX	
DE	Octopus minor; brain; cardiotoxic; cardiant; mollusca; drug;
KW	neural transmission system; agricultural chemical.
KW	
XX	
OS	Octopus minor.
XX	
FT	Key Location/Qualifiers
FT	Misc-difference 2
FT	/label= Ser, Phe
FT	/note= "optionally D-form residue"
FT	Misc-difference 3
FT	/label= Trp, Gly
XX	
PN	JP2000344795-A.
XX	
PD	12-DEC-2000.
XX	
PF	04-JUN-1999; 99JP-0157700.
XX	
PR	04-JUN-1999; 99JP-0157700.
XX	
PA	(SUNR) SUNTORY LTD.
XX	
DR	WPI; 2001-268264/28.
XX	

PT New neuropeptides from brain of Octopus minor are used as cardiotonics
 PS
 XX
 XX
 CC Claim 1; Page 2; 10pp; Japanese.

XX The present sequence represents a cardiotoxic peptide (I) derived from
 CC the brain of Octopus minor. Cardiotoxic peptides of Mollusca of formula
 CC H-Gly-X-Y-Asp-OH (I), where X = L-Ser, D-Ser, L-Phe or D-Phe and
 CC Y = Trp or Gly, obtained from the brain of Octopus minor, can be
 CC used for drugs and agricultural chemicals. (I) have cardiotoxic activity.
 CC (I) can be used as cardiotoxic agents and agricultural chemicals. The
 CC peptides are useful as biochemical reagents for neural transmission.
 CC system and structure and activity relationship in the molecular level
 CC leading to development of drugs and agricultural chemicals.

XX Sequence 4 AA;

Query Match 15.4%; Score 2; DB 22; Length 4;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
 DB 1 gxxd 4

RESULT 2
 AAR75127
 ID AAR75127 standard; Peptide; 9 AA.

XX AAR75127;

XX 15-MAY-1996 (first entry)

XX Putative metal ion binding domain consensus core sequence.

DE Core sequence; metal ion binding domain; consensus;
 DE glycosyl-phosphatidylinositol; phospholipase-D; bovine; liver;
 KW integrin-alpha; EF-hand motif; calcium binding protein;
 KW signal peptide; recombinant protein; anchor; vector; cloning;
 KW fusion protein; protein secretion; purification.

XX Bos taurus.

XX Key Location/Qualifiers

FT Misc-difference 2 /note= "Amino acid not specified"
 FT Misc-difference 4 /note= "Amino acid not specified"
 FT Misc-difference 7 /note= "Amino acid not specified"
 FT Misc-difference 8 /note= "Amino acid not specified"
 FT Misc-difference 8 /note= "Amino acid not specified"

XX US5418147-A.

XX 23-MAY-1995.

XX 27-SEP-1990; 90US-0588896.

XX 31-MAR-1992; 92US-0860825.

XX 27-SEP-1990; 90US-0588896.

XX (HOFF) HOFFMANN LA ROCHE INC.

XX Huang K, Kochan JP, Li SH, Pan YE, Scallan BJ;

PI Tsang TCH;

XX WPI; 1995-199737/26.

XX New glycosyl-phosphatidylinositol-specific phospholipase D gene

PT used partic. for producing secreted recombinant proteins from

PT GPI-anchored proteins expressed by cells
 PS
 XX
 XX Example 2; Column 21; 54pp; English.

XX The sequence represents a consensus core sequence derived from 4
 CC regions of internal similarity in bovine liver
 CC glycosyl-phosphatidylinositol (GPI)-specific phospholipase-D
 CC (AAR75105). The regions show homology with metal ion binding domains
 CC of integrin-alpha subunits, and share an Asp-rich core sequence
 CC flanked by short conserved segments unique to integrins. The core
 CC sequence matches the EF-hand consensus motif characteristic of
 CC calcium and magnesium binding proteins. The presence of these
 CC domains is consistent with calcium requirement of the
 CC phospholipase-D. A gene encoding this type of enzyme may be
 CC connected to a heterologous gene linked at the 3'-end to a C-terminal
 CC signal peptide (e.g. AAR75129 or AAR75130) gene to form a GPI-anchored
 CC molecule, where the C-terminal peptide directs attachment of a GPI
 CC anchor to the heterologous protein. The resulting construct may be
 CC inserted in vector for expression in a host cell. The enzyme
 CC selectively hydrolyses the inositol-phosphate linkage of GPI-anchored
 CC proteins, GPI lipids and related molecules, resulting in direct
 CC secretion into the medium and simple recombinant protein purification.

XX Sequence 9 AA;

Query Match 15.4%; Score 2; DB 16; Length 9;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXND 5
 DB 6 gxxd 9

RESULT 3

AAR86325

ID AAR86325 standard; Peptide; 11 AA.

XX AAR86325;

XX 17-APR-1996 (first entry)

XX [3H]BANA-CKK-8s (affinity ligand specific for anti-CKK antibody).

XX photolytically cleavable affinity ligand; cholecystokinin;
 KW anti-CKK antibody ligand; carrier-binding; substance-binding;
 KW purification; affinity chromatography.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "[3H]4-(Biotin-epsilon-NH(CH2)5CO-oxymethyl)-3-
 FT nitrobenzoyl-Gly; (comprises carrier-binding site,
 FT spacer mol., photolytically cleavable linkage)"

FT Modified-site 2 /note= "Orn-propionyl[3H]"

FT Modified-site 3 /label= OTHER

FT Modified-site 5 /note= "NH(CH2)5CO, 6-aminohexanoic acid"

FT Modified-site 11 /label= Tyr(SO3H)

FT Binding-site 4..11 /note= "Phe-NH2"

FT Binding-site /label= substance_binding_site

XX DE4401135-A1.

XX 07-SEP-1995.

PF 17-JAN-1994; 94DE-4401135.
 PR 17-JAN-1994; 94DE-4401135.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA Fahrenheit F, Thiele C;
 XX WPI; 1995-312004/41.
 XX New photolytically cleavable affinity ligands - with photolytically
 PT cleavable linkage between carrier-binding and substance-binding
 PT sites
 XX
 XX Claim 10; Fig 1A; 13pp; German.
 XX Compounds useful as affinity ligands for e.g. purifying substances are
 CC new, which comprise at least one carrier-binding site (CBS) (eg.
 CC biotin), at least one substance-binding site (SBS), and at least one
 CC photolytically cleavable linkage (PCL) between a CBS and SBS. A spacer,
 CC esp. a 6-aminohexanoic acid gp. may be located between the CBS and SBS
 CC and the PCL is an o-nitrobenzyl ester linkage. AAR66325 peptide
 CC ([3H]-BANA -CCK-8s) is an affinity ligand for anti-CKK (cholecystokinin)
 CC antibodies. Biologically active substances can be isolated in pure form
 CC without damage to the substances by an "affinity chromatography" process
 CC in which the cpd. is incubated with a soln. of the target substance to
 CC form a complex; the complex is bound to a carrier through the CBS; the
 CC PCL is cleaved by irradiation to release a complex comprising the
 CC substance and a fragment of the cpd.; and the substance is recovered from
 CC this complex, e.g. by gel filtration, dialysis or ultrafiltration.
 XX
 SQ Sequence 11 AA;

Query Match 15.4%; Score 2; DB 16; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GXXD 5
 IIII
 Db 1 gxxd 4

RESULT 4
 AAY45210
 ID AAY45210 standard; peptide; 11 AA.
 XX
 AC AAY45210;
 XX
 DT 06-JAN-2000 (first entry)
 XX
 DE Neuroprotective calcium binding peptide SEQ ID NO:15.
 XX
 KW Neuroprotective; calcium binding; stroke; neurodegenerative disease;
 KW blood-brain barrier; cerebral ischaemia; Alzheimer's disease;
 KW memory deficit; aging.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 2
 FT /label= Asp, Asn, Thr, Glu
 FT Misc-difference 3
 FT /note= "any amino acid"
 FT
 XX WO9940112-A1.
 PN
 XX
 XX 12-AUG-1999.
 XX
 XX 28-JAN-1999; 99WO-US01786.
 PF
 XX 10-FEB-1998; 98US-0021247.
 PR
 XX

PA (NEUR-) NEUROMEDICA INC.
 XX Shashoua VE;
 XX WPI; 1999-610582/52.
 DR
 XX
 PT Neuroprotective peptides, which bind calcium, are useful for treating
 PT stroke and other neurodegenerative diseases
 XX
 XX Claim 14; Page 56; 65pp; English.

XX The present invention describes a composition comprising an isolated
 CC peptide, which comprises the amino acid sequence (I) or (Ia):
 CC X1-XX3-XX5-XX7-XX9-XXX12 (I); X5-X6-X7-X8-X9-X10-X11-X12 (Ia); where
 CC X1 = Asp, Gln, Gly or Tyr; X = any amino acid; X3 = Asp, Asn, Thr or Glu;
 CC X5 = Asp, Ser, Gly, Asn or Leu; X7 = Ala, Asp, Phe, Lys, Thr, Tyr, Arg,
 CC Val, Cys or Ser; X9 = Asp, Glu, Gly, Ser, Thr, Met or Asn; and X12 =
 CC Glu, Gln, Ala, Leu or Asn. (I) and (Ia) are neuroprotective calcium
 CC binding peptides. (I) is used to treat a condition characterized by
 CC cerebral ischaemia. (I) reduces the neurotoxic effect of cerebral
 CC ischaemia. (I) is used to increase neuronal cell AP-1 or NF-IL6
 CC transcription factor activity. The peptides are also useful for binding
 CC calcium. The peptide can be conjugated with a compound which facilitates
 CC transport across the blood brain barrier into the brain or it can be
 CC administered with a compound that increases transport across the blood
 CC brain barrier. Molecules that protect neurons against the ischaemic
 CC effects of stroke will also be useful for treating Alzheimer's disease,
 CC as well as the memory deficits that are characteristic of the aging
 CC process. The present sequence represents a specifically claimed example
 CC of a neuroprotective calcium binding peptide from the present invention.
 XX
 SQ Sequence 11 AA;

Query Match 15.4%; Score 2; DB 20; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GXXD 5
 IIII
 Db 1 gxxd 4

RESULT 5
 AAR24843
 ID AAR24843 standard; Protein; 12 AA.
 XX
 AC AAR24843;
 XX
 DT 08-DEC-1992 (first entry)
 XX
 DE Weight regulating peptide 26.
 XX
 KW Amphetamine; appetite suppressor.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 4
 FT /label= GLY, ALA, VAL, LEU, SER, THR, CYS, MET,
 FT ASP, GLU, ASN, GLN, LYS, HIS, ARG, PHE,
 FT TYR, TRP, PRO, OTHER
 FT /note= "cystine, hydroxylysine, hydroxyproline"
 FT Misc-difference 8
 FT /label= GLY, ALA, VAL, LEU, SER, THR, CYS, MET,
 FT ASP, GLU, ASN, GLN, LYS, HIS, ARG, PHE,
 FT TYR, TRP, PRO, OTHER
 FT /note= "cystine, hydroxylysine, hydroxyproline"
 FT Misc-difference 9
 FT /label= GLY, ALA, VAL, LEU, SER, THR, CYS, MET,
 FT ASP, GLU, ASN, GLN, LYS, HIS, ARG, PHE,
 FT TYR, TRP, PRO, OTHER
 FT /note= "cystine, hydroxylysine, hydroxyproline"
 FT

XX PN WO9209296-A.
 XX PD 11-JUN-1992.
 XX PF 20-NOV-1991; 91WO-US08497.
 XX PR 21-NOV-1990; 90US-0616910.
 XX PA (GEOU) UNIV GEORGETOWN.
 XX PT Fleming PJ, Kent UM;
 XX DR WPI; 1992-216791/26.
 XX PT New dodeca-peptide cpds. - used for regulating wt. gain in
 XX PT mammals or for producing antibodies for attenuating such effects
 XX PS Disclosure; Page 10; 34pp; English.
 XX CC The sequences given in AAR24818-61 are new peptides which comprise at
 XX CC least 6 amino acids from the sequence given in AAR24814. The
 XX CC remaining amino acids are each Gly, Ala, Val, Leu, Ser, Thr, Cys,
 XX CC Cysteine, Met, Asp, Glu, Asn, Gln, Lys, hydroxylysine, His, Arg, Phe,
 XX CC Tyr, Trp, Pro or hydroxyproline. These peptides used for the
 XX CC regulation of weight gain in mammals and can be used instead of
 XX CC amphetamine, which is largely used as an appetite suppressor. These
 XX CC peptides can also be used to prepare antibodies. Such antibodies can
 XX CC be used to attenuate the effect of the peptides in a host or to detect,
 XX CC quantify or purify the peptides.
 XX SQ Sequence 12 AA:
 Query Match 15.4%; Score 2; DB 13; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GXXD 5
 Db ||||
 7 gxxd 10
 RESULT 6
 AAW72848
 ID AAW72848 standard; Peptide; 14 AA.
 XX AC AAW72848;
 XX DT 01-MAR-1999 (first entry)
 XX DE Uridine binding motif (consensus).
 XX DE Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
 XX KW amplification; sequencing; replication.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Misc-difference 1 /note= "any residue"
 FT Misc-difference 3 /note= "hydrophobic residue"
 FT Misc-difference 4 /note= "hydrophobic residue"
 FT Misc-difference 6 /note= "any residue"
 FT Misc-difference 7 /note= "any residue"
 FT Misc-difference 8 /note= "any residue"
 FT Misc-difference 9 /note= "hydrophobic residue"
 FT Misc-difference 9 /note= "any residue"

FT Misc-difference 12 /note= "hydrophobic residue"
 FT Misc-difference 14 /note= "hydrophobic residue"
 XX PN WO9842860-A1.
 XX PD 01-OCT-1998.
 XX PF 20-MAR-1998; 98WO-US05497.
 XX PR 24-OCT-1997; 97US-0957709.
 XX PR 21-MAR-1997; 97US-0822774.
 XX PA (STRA-) STRATAGENE.
 XX PT Hansen CJ, Hogrefe H;
 XX DR WPI; 1998-542284/46.
 XX CC Polymerase enhancing factor proteins, extracts and complexes -
 XX CC improve the polymerisation activity of nucleic acid polymerases, for
 XX CC use in amplification, sequencing and replication
 XX PS Claim 71; Page 47; 161pp; English.
 XX CC This is a consensus sequence of uridine-binding motifs of dUTPases
 XX CC and dCTP deaminases of Pyrococcus furiosus, Methanococcus
 XX CC jannaschii, Desulfurolobus ambivalens, Escherichia coli, yeast, human
 XX CC and herpesvirus (see AAW72850-57). A claimed method of enhancing a
 XX CC nucleic acid polymerase reaction comprises performing the reaction
 XX CC in the presence of one or more of the following: a polymerase
 XX CC enhancing factor (PEF, e.g. P45 of P. furiosus, see AAW72847), a
 XX CC dUTPase, a protein that turns-over dUTP and a protein selected from
 XX CC the sequences provided in AAW72848-57. A claimed protein having PEF
 XX CC activity comprises one or more of the sequences given in AAW72848-57.
 XX CC The invention provides novel extracts, proteins and complexes that
 XX CC improve the polymerisation activity of nucleic acid polymerases.
 XX CC Also included are methods for identifying compositions with
 XX CC polymerase enhancing activity, for purifying and using these
 XX CC compositions, and specific extracts, proteins and complexes that
 XX CC function to enhance polymerase activity. Kits are provided for
 XX CC replicating nucleic acids, for site-directed mutagenesis, for
 XX CC nucleic acid sequencing or for amplification (preferably PCR or
 XX CC RT-PCR).
 XX SQ Sequence 14 AA:
 Query Match 15.4%; Score 2; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GXXD 5
 Db ||||
 2 gxxd 5
 RESULT 7
 AAG62873
 ID AAG62873 standard; peptide; 14 AA.
 XX AC AAG62873;
 XX DT 17-SEP-2001 (first entry)
 XX DE Amino acid sequence of a conserved region of Fam35-lactases.
 XX KW Fam35-lactase; fungus; foodstuff manufacture; lactose intolerance.
 XX OS Unidentified.
 XX FH Key Location/Qualifiers

FT Misc-difference 4 /label= Ala, Gly
 FT Misc-difference 6 /label= Ala, Ser
 FT Misc-difference 7 /label= Phe, Tyr
 FT
 PN WO200149878-A1.
 XX
 XX 12-JUL-2001.
 XX
 XX 14-DEC-2000; 2000WO-DK00693.
 XX
 XX 30-DEC-1999; 99DK-0001888.
 PR 13-MAR-2000; 2000DK-0000397.
 PR 13-OCT-2000; 2000DK-0001529.
 XX
 XX (NOVO) NOVOZYMES AS.
 PA
 XX Schnorr K, Lange L, Lassen SF;
 PI WPI; 2001-441894/47.
 XX
 DR
 XX

XX Screening for DNA encoding enzyme of interest, involves obtaining DNA
 PT from microorganism, selecting PCR primers, performing PCR and screening
 FT PCR products, or selecting a probe for use in Southern blot -
 XX
 XX Claim 19; Page 63; 123pp; English.

CC AAG62871-77 represent conserved regions of Fam35-lactases. The
 CC specification describes a method for obtaining new Fam35-lactases
 CC of fungal origin. Primers designed from highly conserved regions of
 CC Fam35-lactases are used in the method of the invention. The
 CC Fam35-lactases are useful in the manufacture or processing of
 CC foodstuffs or feeds fit for consumption by lactose intolerant humans
 CC or animals. Fam35-lactases are also useful for treating lactose
 CC intolerance in mammals.
 XX
 XX Sequence 14 AA;

Query Match 15.4%; Score 2; DB 22; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
 DB 5 gxnd 8

RESULT 8
 AAY42757
 ID AAY42757 standard; peptide; 15 AA.
 XX
 XX AAY42757;
 AC
 XX
 XX 20-DEC-1999 (first entry)
 DT
 XX

DE Wheat amyloplast ADP-glucose transporter peptide #5.

XX Starch biosynthesis; amyloplast; ADP-glucose; transport; import;
 KW amylopectin; amylose; branching; chemical structure; transgenic plant;
 KW optimisation; industrial applications.
 XX
 XX Triticum aestivum.

OS
 XX Key Location/Qualifiers
 FH Misc-difference 7 /label= Xaa
 FT /note= "Xaa = any amino acid"
 FT Misc-difference 8 /label= Xaa
 FT /note= "Xaa = any amino acid"
 FT
 FT

FT Misc-difference 10 /label= Xaa
 FT /note= "Xaa = any amino acid"
 FT Misc-difference 11 /label= Xaa
 FT /note= "Xaa = any amino acid"
 FT Misc-difference 12 /label= Xaa
 FT /note= "Xaa = any amino acid"
 XX
 XX
 PN WO9947682-A1.
 XX
 XX 23-SEP-1999.
 XX
 XX 19-MAR-1999; 99WO-GB00728.
 XX
 XX 20-MAR-1998; 98GB-0005939.
 XX
 XX (UYMA-) UNIV VICTORIA MANCHESTER.
 PA
 XX Emes MJ, Tetlow IJ, Bowsheer CG;
 PI WPI; 1999-590977/50.
 XX

XX New transporter complex protein useful for modulating starch content in
 PT plants, especially useful in food production -
 PT
 XX
 XX Claim 1; Page 3; 28pp; English.

XX This sequence represents a wheat amyloplast ADP-glucose transporter
 CC peptide, #5. The wheat amyloplast ADP-glucose transporter is associated
 CC with the amyloplast membrane and comprises at least two proteins; this
 CC sequence, along with peptides #4-#7 (AAY42756-Y42759) represent
 CC sequences within one of these proteins (AAY42760). Peptides #1-#3
 CC (AAY42753-Y42755) are components of a second protein. The sugar
 CC nucleotide ADP-glucose is the immediate substrate for starch synthesis,
 CC which occurs in the amyloplast; however, ADP-glucose is mainly
 CC synthesised outside the amyloplast in the cytoplasm. The ADP-glucose
 CC transporter is responsible for the import of ADP-glucose into the
 CC amyloplast and therefore plays a pivotal role in the regulation of starch
 CC synthesis. The transporter not only influences starch yield, but also
 CC quality as the starch syntheses involved in amylose and amylopectin
 CC synthesis have different affinities for ADP-glucose. Variations in the
 CC chemical structure of starch are determined by the ratio of amylose to
 CC amylopectin, and by the degree of branching in amylopectin in the starch
 CC polymer. These variations can significantly alter the properties of
 CC starch. The ADP-glucose transporter complex is useful for generating
 CC transgenic plants in which the starch quality can be optimised for
 CC industrial applications in which starch is used. For example, transgenic
 CC plants which have an increased amylose content in starch are useful for
 CC production of starch with increased viscosity and gel strength, which
 CC prevents baked food going stale so quickly. Conversely, an increased
 CC amylopectin content in the starch produces waxy starch useful as
 CC thickening agents in food and coatings.
 XX
 XX Sequence 15 AA;

Query Match 15.4%; Score 2; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DXXXG 9
 DB 9 dxxxg 13

RESULT 9
 AAY39335
 ID AAY39335 standard; peptide; 15 AA.
 XX
 XX AAY39335;
 AC
 XX

DT 01-DEC-1999 (first entry)
 XX ADP glucose transporter peptide 5.
 XX
 XX ADP glucose transporter; transform plant cell; wheat; starch production;
 KW waxy starch; thickening agent; food; coating; increased viscosity; stale;
 KW gel strength; baked food.
 XX
 XX Triticum aestivum.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 7 /note= "Any amino acid"
 FT Misc-difference 8 /note= "Any amino acid"
 FT Misc-difference 10 /note= "Any amino acid"
 FT Misc-difference 11 /note= "Any amino acid"
 FT Misc-difference 12 /note= "Any amino acid"
 FT
 FT WO9947681-A1.
 XX
 XX 23-SEP-1999.
 XX
 XX 19-MAR-1999; 99WO-GB00727.
 XX
 XX 20-MAR-1998; 98GB-0005939.
 XX (UYMA-) UNIV VICTORIA MANCHESTER.
 XX
 XX Emes MJ, Tetlow IJ, Bowsher CG;
 XX WPI; 1999-571841/48.
 DR
 XX ADP glucose transporter protein used for modifying plant starch
 FT production
 FT
 XX Claim 1; Page 15; 26pp; English.
 PS
 CC Peptides AAY39331-Y39337 are fragments of an ADP glucose transporter
 CC protein. The protein contains at least one of the peptide sequences, and
 CC is capable of ADP glucose transport. A DNA molecule encoding an ADP
 CC glucose transporter protein can be used to transform plant cells. The
 CC ADP glucose transporter protein can be used to regulate starch
 CC production from a plant. The plants can then be used to produce waxy
 CC starches that can be used as thickening agents in food and coatings.
 CC Alternatively the plants can be used to produce a starch with increased
 CC viscosity and gel strength, which can be incorporated in baked food which
 CC stays fresh for a longer length of time.
 XX
 SQ Sequence 15 AA;
 Query Match 15.4%; Score 2; DB 20; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 DXXGX 9
 Db 9 dxxxg 13
 |||||
 RESULT 10
 AAW82117
 ID AAW82117 standard; peptide; 19 AA.
 XX
 AC AAW82117;
 XX
 DT 18-FEB-1999 (first entry)
 XX
 XX Fluorogenic protease Indicator CPP32 substrate #6.

XX Protease activity; fluorophore: detection; fluorogenic; cellular uptake;
 KW conformation change.
 XX Synthetic.
 XX Key Location/Qualifiers
 FT Modified-site 3 /label= Alb
 FT /note= "alpha-aminoisobutyric acid, labelled as
 FT amino acid B in the specification"
 FT Modified-site 4 /note= "epsilon-aminocaproic acid, labelled as
 FT amino acid J in the specification"
 FT Modified-site 5 /note= "epsilon-aminocaproic acid, labelled as
 FT amino acid J in the specification"
 FT Modified-site 14 /note= "epsilon-aminocaproic acid, labelled as
 FT amino acid J in the specification"
 FT Modified-site 15 /note= "epsilon-aminocaproic acid, labelled as
 FT amino acid J in the specification"
 FT
 FT WO9837226-A1.
 XX
 XX 27-AUG-1998.
 XX
 XX 20-FEB-1998; 98WO-US03000.
 XX
 XX 20-FEB-1997; 97US-0802981.
 XX (ONCO-) ONCOIMMUNIN INC.
 XX
 XX Komoriya A, Packard BS;
 XX WPI; 1998-467579/40.
 DR
 XX New fluorogenic compositions - containing 2 fluorophores separated
 FT by a peptide comprising a protease binding site, used for detecting
 FT protease activity in samples.
 XX
 XX Disclosure; Page 24; 90pp; English.
 PS
 CC AAW82023-W82240 are peptides used in the construction of a fluorogenic
 CC composition which is used for the detection of protease activity in
 CC biological samples. The products can be used for the detection of
 CC conformational changes in nucleic acids, oligosaccharides,
 CC polysaccharides, proteins, peptides, lipids, phospholipids, glycolipids,
 CC glycoproteins, steroids or polymers. In addition, attachment of a
 CC hydrophobic group to a molecule can be used to enhance uptake by cells.
 CC The composition is composed of P = peptide comprising a protease binding
 CC site for the protease, F1, F2 peptides = fluorophores where F1 is
 CC attached to the amino terminal amino acid and F2 is attached to the
 CC carboxyl terminal amino acid and S1, S2 peptides = when present, are
 CC peptide spacers where S1, when present, is attached to the amino terminal
 CC acid, and S2, when present, is attached to the carboxyl terminal amino
 CC acid.
 XX
 SQ Sequence 19 AA;
 Query Match 15.4%; Score 2; DB 19; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 DXXGX 9
 Db 2 dxxxg 6
 |||||
 RESULT 11
 AAW82125

AAW82125 standard; peptide; 20 AA.
AAW82125;
18-FEB-1999 (first entry)
Fluorogenic protease indicator CPP32 substrate #14.
Protease activity; fluorophore; detection; fluorogenic; cellular uptake;
conformation change.
Synthetic.
Key Location/Qualifiers
Modified-site 3 /label= Aib
/note= "alpha-aminoisobutyric acid, labelled as
amino acid B in the specification"
Modified-site 4 /note= "epsilon-aminocaproic acid, labelled as
amino acid J in the specification"
Modified-site 6 /note= "tetrahydroisoquinoline-3-carboxylic acid,
labelled as amino acid O in the specification.
Can be L- or D- form residue"
Modified-site 7 /note= "tetrahydroisoquinoline-3-carboxylic acid,
labelled as amino acid O in the specification.
Can be L- or D- form residue"
Modified-site 16 /note= "epsilon-aminocaproic acid, labelled as
amino acid J in the specification"
WO9837226-A1.
27-AUG-1998.
20-FEB-1998; 98WO-US03000.
20-FEB-1997; 97US-0802981.
(ONCO-) ONCOIMMUNIN INC.
Komoriya A, Packard BS;
WPI; 1998-467579/40.
New fluorogenic compositions - containing 2 fluorophores separated
by a peptide comprising a protease binding site, used for detecting
protease activity in samples.
Disclosure; Page 25; 90pp; English.
AAW82023-W82240 are peptides used in the construction of a fluorogenic
composition which is used for the detection of protease activity in
biological samples. The products can be used for the detection of
conformation changes in nucleic acids, oligosaccharides,
polysaccharides, proteins, peptides, lipids, phospholipids, glycolipids,
glycoproteins, steroids or polymers. In addition, attachment of a
hydrophobic group to a molecule can be used to enhance uptake by cells.
The composition is composed of P = peptide comprising a protease binding
site for the protease, F1, F2 peptides = fluorophores where F1 is
attached to the amino terminal amino acid and F2 is attached to the
carboxyl terminal amino acid and S1, S2 peptides = when present, are
peptide spacers where S1, when present, is attached to the amino terminal
acid, and S2, when present, is attached to the carboxyl terminal amino
acid.
Sequence 20 AA;
Query Match 15.4%; Score 2; DB 19; Length 20;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 DXXXG 9
Db 14 dxxxg 18

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GXXD 5
Db 5 gxxd 8
RESULT 12
AAG76715
ID AAG76715 standard; Protein; 20 AA.
XX
AC AAG76715;
XX
XX 03-SEP-2001 (first entry)
DE Human colon cancer antigen protein SEQ ID NO:7479.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
PF
XX 29-SEP-1999; 99US-0157137.
PR
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI
XX WPI; 2001-235357/24.
DR
DR N-PSDB; AAH36120.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 8904-8905; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB7789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 20 AA;
Query Match 15.4%; Score 2; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 DXXXG 9
Db 14 dxxxg 18

```

RESULT 13
AAG73106
ID AAG73106 standard; Peptide; 20 AA.
XX
AC AAG73106;
XX
AC AAG73106;
XX
DT 14-AUG-2001 (first entry)
XX
DE Protease binding site #40.
XX
KW Protease detection; peptide cleavage; enzyme activity; fluorogenic;
KW viral infection; cancer metastasis; emphysema; arthritis;
KW thrombosis; haemophilia.
XX
OS Synthetic.
XX
PH KEY Location/Qualifiers
FT Modified-site 3 /label= Alb
FT /note= "2-aminoisobutyric acid"
FT Modified-site 6..7 /label= OTHER
FT /note= "tetrahydroisouquinoline-3-carboxylic acid,
FT D-form residue"
XX
PN WO200118238-A1.
XX
PD 15-MAR-2001.
XX
PF 11-SEP-2000; 2000WO-US24882.
XX
PR 10-SEP-1999; 99US-0394019.
XX
PA (ONCO-) ONCOIMMUNIN INC.
XX
PI Komoriya A, Packard BS;
XX
DR WPI; 2001-389573/41.
XX
XX
XX New fluorogenic compositions whose fluorescence level increases in the
XX presence of active proteases, useful for detecting and localizing
XX protease activity in biological samples, particularly in frozen tissue
XX samples.
XX
XX Disclosure; Page 24; 86pp; English.
XX
XX The present invention describes fluorogenic compositions which can be
XX used for the detection of protease activity. This can be useful as an
XX indicator of viral infection, cancer metastasis, haemophilia, emphysema,
XX thrombosis and arthritis. The fluorogenic compositions comprise a
XX peptide, a peptide spacer and a donor and an acceptor fluorophore. The
XX peptide is cleaved by a protease and the fluorophores can then be
XX detected. The present sequence is one of the peptides described in the
XX exemplification of the invention.
XX
SQ Sequence 20 AA:

Query Match 15.4%; Score 2; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GXXD 5
Db 5 gxxd 8
|||||

RESULT 14
AAB81828
ID AAB81828 standard; Peptide; 20 AA.
XX

```

```

AC
XX
DT
XX
DE
XX
DE Human endozepine-like ENDO proteins consensus peptide SEQ ID NO: 39.
XX
KW Human; endozepine-like protein; ENDO1; ENDO2; ENDO3; ENDO4; ENDO5;
KW ENDO6; ENDO7; ENDO8; ENDO9; ENDO10; obesity; diabetes; dyslipidaemia;
KW metabolic disorder; infection; anorexia; cachexia; syndrome X.
XX
OS Homo sapiens.
XX
PH KEY Location/Qualifiers
FT misc_difference 6 /label= Asp, Pro
FT misc_difference 8 /label= Lys, Arg, Asn
FT misc_difference 9 /label= Ala, Ile, Thr, Val, Phe, Leu, Met
FT misc_difference 10 /label= Xaa
FT /note= "Xaa=unknown"
FT misc_difference 11 /label= Lys, Arg
FT misc_difference 14 /label= Ala, Ile, Thr, Phe, Leu, Met
FT misc_difference 15 /label= Met, Leu, Phe, Val, Thr, Ile, Ala, Trp
FT misc_difference 17 /label= Lys, Met, Phe, Val, Thr, Ile, Ala, Pro
FT misc_difference 19 /label= Xaa
FT /note= "Xaa=unknown"
FT misc_difference 20 /label= Lys, Arg
XX
PN WO200125436-A2.
XX
PD 12-APR-2001.
XX
XX
XX 05-OCT-2000; 2000WO-US41077.
XX
XX 05-OCT-1999; 99US-0157786.
XX 09-NOV-1999; 99US-0164164.
XX 04-JAN-2000; 2000US-0174505.
XX 22-FEB-2000; 2000US-0183859.
XX 20-MAR-2000; 2000US-0190740.
XX 22-MAR-2000; 2000US-0191133.
XX 19-MAY-2000; 2000US-0206006.
XX 30-JUN-2000; 2000US-0215684.
XX 20-JUL-2000; 2000US-0219490.
XX 22-AUG-2000; 2000US-0227072.
XX
PA (CURA-) CURAGEN CORP.
XX
XX Prayaga SK, Shimkets RA, Majumder K, Eisen A, Vernet C;
XX Spaderna SK;
XX
XX WPI; 2001-266310/27.
XX
XX Novel human endozepine-like polypeptide, polynucleotide for diagnosis,
XX prevention, treatment of diabetes, metabolic disturbances associated
XX with obesity, anorexia, cancer, dyslipidaemia and for identifying
XX modulators.
XX
XX Claim 1; Page 115; 136pp; English.
XX
XX The present invention provides the protein and coding sequences of 10
XX endozepine-like proteins, designated ENDO1-ENDO10. The sequences are
XX useful in the treatment of various diseases, including obesity, metabolic
XX disorders, diabetes, infections, anorexia, syndrome X, cancer-associated
XX cachexia and dyslipidaemias. The present sequence is a protein fragment
XX of the invention.

```



```
XX SQ Sequence 20 AA;
Query Match 15.4%; Score 2; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GXXD 5
Db 13 gxxd 16
||||
RESULT 15
AAB81831
ID AAB81831 standard; Peptide; 20 AA.
XX AC AAB81831;
XX 05-JUN-2001 (first entry)
XX Human endozepine-like ENDO proteins consensus peptide SEQ ID NO: 42.
XX Human; endozepine-like protein; ENDO1; ENDO2; ENDO3; ENDO4; ENDO5;
KW ENDO6; ENDO7; ENDO8; ENDO9; ENDO10; obesity; diabetes; dyslipidaemia;
KW metabolic disorder; infection; anorexia; cachexia; syndrome X.
XX Homo sapiens.
XX Key Location/Qualifiers
FH misc_difference 6 /label= Asp, Asn, Pro
FT misc_difference 7 /label= Ile, Cys
FT misc_difference 9 /label= Thr, Ile, Met, Leu
FT misc_difference 10 /label= Xaa
FT misc_difference 11 /note= "Xaa=unknown"
FT misc_difference 14 /label= Arg, Lys
FT misc_difference 14 /label= Met, Xaa
FT misc_difference 15 /note= "Xaa=unknown"
FT misc_difference 15 /label= Xaa
FT misc_difference 17 /note= "Xaa=unknown"
FT misc_difference 17 /label= Phe, Leu
FT misc_difference 18 /label= Xaa
FT misc_difference 20 /note= "Xaa=unknown"
FT misc_difference 20 /label= Lys, Arg
XX WO200125436-A2.
XX 12-APR-2001.
XX 05-OCT-2000; 2000WO-US41077.
XX 05-OCT-1999; 99US-0157786.
XX 09-NOV-1999; 99US-0164164.
XX 04-JAN-2000; 2000US-0174505.
XX 22-FEB-2000; 2000US-0183859.
XX 20-MAR-2000; 2000US-0190740.
XX 22-MAR-2000; 2000US-0191133.
XX 19-MAY-2000; 2000US-0206006.
XX 30-JUN-2000; 2000US-0215684.
XX 20-JUL-2000; 2000US-0219490.
XX 22-AUG-2000; 2000US-0227072.
XX (CURA-) CURAGEN CORP.
Prayaga SK, Shinkets RA, Majumder K, Eisen A, Vernet C;
Spaderna SK;
WPI; 2001-266310/27.
Novel human endozepine-like polypeptide, polynucleotide for diagnosis,
prevention, treatment of diabetes, metabolic disturbances associated
with obesity, anorexia, cancer, dyslipidaemia and for identifying
modulators -
Claim 1; Page 116; 136pp; English.
The present invention provides the protein and coding sequences of 10
endozepine-like proteins, designated ENDO1-ENDO10. The sequences are
useful in the treatment of various diseases, including obesity, metabolic
disorders, diabetes, infections, anorexia, syndrome X, cancer-associated
cachexia and dyslipidaemias. The present sequence is a protein fragment
of the invention.
XX SQ Sequence 20 AA;
Query Match 15.4%; Score 2; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GXXD 5
Db 13 gxxd 16
||||
Search completed: January 31, 2002, 13:18:06
Job time: 167 sec
```

This Page Blank (uspto)

.....

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 10:53:12 ; Search time 101.18 Seconds
(without alignments)
3990.909 Million cell updates/sec

Title: US-08-957-709-70
Perfect score: 471
Sequence: 1 ATGCTTACTTCCAGACTGGAA.....CRAAGAGAAAGAACTCTAG 471

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	471	100.0	471	19	AAV63860 Polymerase enhanc
2	44.8	9.5	936	22	AAF58252 Oligonucleotide D1
3	44.8	9.5	936	22	AAF58254 Oligonucleotide D1
4	44.8	9.5	936	22	AAF58257 Oligonucleotide D1
5	44.8	9.5	936	22	AAF58259 Oligonucleotide D2
6	44.8	9.5	936	22	AAF58262 Oligonucleotide D2
7	44.8	9.5	936	22	AAF58255 Oligonucleotide D1
8	44.8	9.5	3382	14	AAO62300 pDam-L3 plasmid fr
9	44.4	9.4	936	22	AAF58252 Oligonucleotide D1
10	44.4	9.4	936	22	AAF58254 Oligonucleotide D1
11	44.4	9.4	936	22	AAF58257 Oligonucleotide D1

12	44.4	9.4	936	22	AAF58259 Oligonucleotide D2
13	44.4	9.4	936	22	AAF58262 Oligonucleotide D1
14	44.4	9.4	936	22	AAF58255 Oligonucleotide D1
15	44.2	9.4	2006	20	AAV90555 Nucleotide sequenc
16	44.2	9.4	2588	20	AAV90871 Nucleotide sequenc
17	38.6	8.2	89047	22	AAF28547 Genomic fragment #
18	37.4	7.9	9202	8	AAV70608 VISA sheep lentiv
19	35	7.4	3257	17	AAI14352 Nsk2 receptor gene
20	35	7.4	3257	19	AAV38546 Mouse receptor tyr
21	34.8	7.4	570	22	AAI17857 Probe #7790 for ge
22	34.8	7.4	570	22	AAI40941 Probe #9627 used t
23	34.8	7.4	570	22	AAI42828 Probe #11514 used
24	33.6	7.1	405	20	AAV98102 Nucleotide sequenc
25	33.6	7.1	765	20	AAV98101 Nucleotide sequenc
26	33.6	7.1	42235	20	AAV98035 Nucleotide sequenc
27	31.8	6.8	491	20	AAV40086 Gastric cancer ass
28	31.6	6.7	10732	21	AAI10594 Gene encoding a su
29	31.4	6.7	2416	22	AAI60592 Human polynucleoti
30	31.4	6.7	2646	16	AAO2405 Rice mature pullul
31	31.4	6.7	2783	21	AAV96712 Reporter gene SVS1
32	31.4	6.7	2982	16	AAO2400 Pullulanase expres
33	31.4	6.7	2988	16	AAO2399 Rice mature pullul
34	31.4	6.7	3398	22	AAO8421 Human secreted pro
35	31.4	6.7	3398	22	AAO5105 Human secreted pro
36	31.4	6.7	7644	22	AAO5067 Human secreted pro
37	31.4	6.7	9474	19	AAV58053 FIV F14 clone. Fe
38	31.2	6.6	779	22	AAV22679 Human gastric canc
39	31.2	6.6	802	20	AAV40087 Gastric cancer ass
40	31.2	6.6	960	19	AAV59279 Human telomere rep
41	31.2	6.6	1311	19	AAV59280 Altered telomere r
42	31.2	6.6	1317	19	AAV59278 Human telomere rep
43	31.2	6.6	1320	22	AAF63966 Human TRF1 coding
44	31.2	6.6	1581	17	AAI16350 Mouse muscle-local
45	31.2	6.6	1629	18	AAV7489 Telomeric repeat b

ALIGNMENTS

RESULT	1
AAV63860	
ID	AAV63860 standard; DNA; 471 BP.
XX	AC
AC	AAV63860;
DT	01-MAR-1999 (first entry)
XX	Polymerase enhancing factor P45 (dUTPase) component DNA.
DE	Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
XX	Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW	amplification; sequencing; replication; ss.
XX	Pyrococcus furiosus strain DSM 3638.
OS	Pyrococcus furiosus strain DSM 3638.
XX	WO9842860-A1.
PN	01-OCT-1998.
PD	01-OCT-1998.
XX	20-MAR-1998; 98WO-US05497.
FF	20-MAR-1998; 98WO-US05497.
XX	24-OCT-1997; 97US-0957709.
PR	21-MAR-1997; 97US-0822774.
XX	(STRA-) STRATAGENE.
PA	Hansen CJ, Hogrefe H;
XX	WPI; 1998-542284/46.
PI	P-PSDB; AAV72847.
DR	Polymerase enhancing factor proteins, extracts and complexes -
DR	improve the polymerisation activity of nucleic acid polymerases, for
XX	use in amplification, sequencing and replication
PT	

XX Claim 17; Page 43; 16lpp; English.

XX This DNA sequence encodes the P45 component (see AAW72847) of the

XX polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.

CC PCR primers (see AAV63861-64) based on an N-terminal peptide (see

CC AAW72846) of P45 were used to amplify *P. furiosus* genomic DNA, and

CC sequencing was carried out on purified PCR products and plasmid

CC mini-preps. P45 and P40 (see AAW72844) are the predominant components

CC of PEF, which acts to enhance the activity of *P. furiosus* DNA

CC polymerase, thereby providing replicated products of greater

CC length and purity. P45 was identified as a dUTPase, and possesses

CC polymerase enhancing activity. The invention provides novel

CC extracts, proteins and complexes that improve the polymerisation

CC activity of nucleic acid polymerases, as well as DNA constructs and

CC antibodies. Also included are methods for identifying compositions

CC with polymerase enhancing activity, for purifying and using these

CC compositions, and specific extracts, proteins and complexes that

CC function to enhance polymerase activity. Nucleic acid polymerase

CC reactions can be enhanced (claimed) by mixing a nucleic acid

CC template, at least 1 polymerase and a composition having polymerase

CC enhancing activity. Kits are provided for replicating nucleic

CC acids. The kits can be used in site-directed mutagenesis, nucleic

CC acid sequencing or amplification (preferably PCR or RT-PCR).

XX Isolated DNA can be used in the recombinant production of P45.

XX Sequence 471 BP; 156 A; 80 C; 124 G; 111 T; 0 other;

SQ

Query Match 100.0%; Score 471; DB 19; Length 471;

Best Local Similarity 100.0%; Pred. No. 6.2e-143;

Matches 471; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATCTTATAGCCATTTTCTGAAGAA 60

DB 1 atgctacttccagactggaaatcagaaagaaatcttattatagccattttctgaagaa 60

QY 61 TCGCTCCAAACGACAGGTATGACCTCAGAGTGGCAGAGGCTTTTGTAAAGGGAAA 120

DB 61 tcgctccaacgacaggttatgacctcagagtggcagagagctttgttaagggaaa 120

QY 121 TTAATCGAGCTGGAAAGGAAAGTCGTTATTCTCCAAAGGAATACGCCCTTAATC 180

DB 121 ttaatcgactggaaaggaagaaagtcgttattctccaaaggaaatacgccttaatc 180

QY 181 CTAACCCCTCGAGAGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240

DB 181 ctaaccctcgagagataaagtgtgcccgacgatgttatggggatatagaagataaggagc 240

QY 241 ACTTTAGCAAGAGAGGGGTATTGTTCTTTTGTGGTTGACCCAGGATGGGATGTA 300

DB 241 agtttagcaagagaggggtattgttcttttggtttgacccaggatgggatatgagagagga 300

QY 301 AACTTAACACTAATGCTCTACAAATGCCCTCAATGAACCTGCAATTAAGATATGAGAG 360

DB 301 aacttaacactaatgctctacaatgccctcaatgaacctgcaatttaagataagagagag 360

QY 361 AGATTGTGCAGATCGCATTTAAGGCTAGAGGTCCGGCAAGAACCCCTTACAGAGGA 420

DB 361 agattgtgcagatcgcatttaaggctagaggtccggcaagaaccccttacagagga 420

QY 421 AACTATCAGGGGAGCACAGGTAGCGTTTTCACAAAGAGAAAGAACTCTAG 471

DB 421 aactatcaggggagcacaggttagcgTTTTACAAAGAGAAAGAACTCTAG 471

RESULT 2

ID AAF58252/C

XX AAF58252 standard; DNA; 936 BP.

AC AAF58252;

XX

XX 24-APR-2001 (first entry)

DT

XX Oligonucleotide D1835.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

OS WO200107665-A2.

PN 01-FEB-2001.

PD 26-JUL-2000; 2000WO-US20476.

PF 17-JUL-1999; 99US-0145695.

PR 26-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in

PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

PI a single surface

XX Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic

CC acids each containing an electron-transfer group (ETM) having

CC different redox potentials. The invention is used for electronic

CC detection of nucleic acids, especially of substitutions (mismatches)

CC and single-nucleotide polymorphisms, e.g. for genotyping,

CC monitoring gene expression.

XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

SQ

Query Match 9.5%; Score 44.8; DB 22; Length 936;

Best Local Similarity 0.9%; Pred. No. 0.00016;

Matches 4; Conservative 258; Mismatches 190; Indels 0; Gaps 0;

QY 19 AAAATCAGAAAGAAATACTTATAGAGCCTTTTCTGAAGAAATCGCTCAACACGACGCT 78

DB 19 aaaaatcagaaagaaatactttatagagccttttctgaagaaatcgctcaacacgacgct 78

QY 79 TATGACCTCAGAGTGGCGAGAGAGGCTTTTGTGAAGGGAATTAATCGACGTGGAAG 138

DB 79 tatgacctcagagtgggcagagagggcttttgtgaagggaattaatcgacgtggaaag 138

QY 714 WW 655

DB 714 ww 655

QY 139 GAAGGAAAAGTCGTTATTCTCCCAAGGGAATACGCTTAATCCTAACCTCGAGAGGATA 198

DB 139 gaaggaagtcggttattctcccaagggaatacgccttaatcctaacctcgagaggata 198

QY 654 WW 595

DB 654 ww 595

QY 199 AAGTTGCCCGAGAGTGTATGGGGGATATGAAGATAAGAGCAGTGTAGCAAGAGAGGG 258

DB 199 aagttgcccgagagtgtatgggggatatgaagataagagcagtgtagcaagagagg 258

QY 594 WWCWW 535

DB 594 wwcww 535

QY 259 GTTATTGTTCTTTTGTGTTGCTTGACCCAGAGTGGGATGGAACCTTAACACTAATGCTC 318

DB 259 gttattgttcttttgtgttgcttgacccagagtgggatggaaaccttaacactaatgctc 318

QY 534 WW 475

DB 534 ww 475

QY 319 TACAAATGCTCAAAATGAACCTGTTCGAATTAAGATATGAGAGAGAGATTGTGTCAGATCGCA 378

DB 319 tacaaatgctcaaatgaacctgttcgaattaaagatatgagagagagattgtgtcagatcgca 378

QY 474 WW 415

DB 474 ww 415

QY 379 TTTATAAGGCTAGAGGGTCCCGCAAGAAACCTTACAGAGGAAACTATCAGGGGAGCACA 438

DB 379 ttтатаaggctagaggggtcccgcaagaaaccttacagaggaaactatcaggggagcacaca 438

QY 414 WW 355

DB 414 ww 355

QY 439 AGGTTAGCGTTTTCAGAGAGAGAAAGAACTCTA 470

DB 439 aggttagcgTTTTTCAGAGAGAGAAAGAACTCTA 470

Db 354 WWWWWW

RESULT 3

AAF58254/c
ID AAF58254 standard; DNA; 936 BP.

XX AAF58254;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D1875.

DE Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

PS Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 9.5%; Score 44.8; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 0.00016;
Matches 4; Conservative 258; Mismatches 190; Indels 0; Gaps 0;

Qy 19 AAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAAATCCTCCACACGAGGT 78
Db 774 WWWWWW

Qy 79 TATGACCTCAGAGTGGGCGAGAGGCTTTTGTGAAGGGAATTAATCGACGTGAAAG 138
Db 714 WWWWWW

Qy 139 GAAGAAAGTCGTTATTCCTCCAGGAATACGCTTTAATCCTAACCTCGAGAGGATA 198
Db 654 WWWWWW

Qy 199 AAGTTCGCCAGGATGTTATGGGGATATGAAGATAAGACGAGTTCAGAGAGAGGG 258
Db 594 WWCWWWWW

Qy 259 GTATTGCTCTTTTGGCTTGCGTCCAGGATGGGATGGAACCTTAACACTAATGCTC 318
Db 534 WWWWWW

Qy 534 WWWWWW

Qy 319 TACAATGCTCAATCAACCTGTGCAATTAAGATATGAGAGAGATTTCTGCAGATCGCA 378
Db 474 WWWWWW
Qy 379 TTTATAAGCTTAGAGGTCGGCAAGAACCTTTACAGAGGAAACTATCAGGGGAGCACA 438
Db 414 WWWWWW
Qy 439 AGGTTAGCCTTTTCAAGAGAGAAAGAACTCTA 470
Db 354 WWWWWW

RESULT 4

AAF58257/c
ID AAF58257 standard; DNA; 936 BP.

XX AAF58257;

XX 24-APR-2001 (first entry)

XX Oligonucleotide D1954.

XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX Synthetic.

XX WO200107665-A2.

XX 01-FEB-2001.

XX 26-JUL-2000; 2000WO-US20476.

XX 26-JUL-1999; 99US-0145695.

XX 17-MAR-2000; 2000US-0190259.

XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX Umek RM;

XX WPI; 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface

PS Example 6; Page 127; 159pp; English.

XX The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 9.5%; Score 44.8; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 0.00016;
Matches 4; Conservative 258; Mismatches 190; Indels 0; Gaps 0;

Qy 19 AAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAAATCCTCCACACGAGGT 78
Db 774 WWWWWW

Qy 79 TATGACCTCAGAGTGGGCGAGAGGCTTTTGTGAAGGGAATTAATCGACGTGAAAG 138
Db 714 WWWWWW

Qy 139 GAAGAAAGTCGTTATTCCTCCAGGAATACGCTTTAATCCTAACCTCGAGAGGATA 198
Db 534 WWWWWW

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 10:22:37 ; Search time 1332.95 Seconds
(without alignments)
5829.305 Million cell updates/sec

Title: US-08-957-709-70
Perfect score: 471
Sequence: 1 ATGCTACTTCAGACTGGAA.....CAAGAGAAAGAACTCTAG 471

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_om:*

20: em_or:*

21: em_ov:*

22: em_pat:*

23: em_ph:*

24: em_pl:*

25: em_ro:*

26: em_sts:*

27: em_sy:*

28: em_un:*

29: em_vl:*

30: em_htgo_hum:*

31: em_htgo_inv:*

32: em_htgo_rod:*

33: em_htg_hum:*

34: em_htg_inv:*

35: em_htg_rod:*

36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	374	79.4	253505	1	AP000007	AP000007 Pyrococcus
2	280.8	59.6	265118	1	CNSPAX06	AJ248288 Pyrococcus
3	280.8	59.6	265118	6	AX041922	AX041922 Pyrococcus
4	90.8	19.3	600	14	AF022221	AF022221 Archaeal
5	54.2	11.5	7218	6	I66494	I66494 Sequence 14
6	49	10.4	11390	1	AE000554	AE000554 Helicobac
7	47.4	10.1	239340	1	AP000993	AP000993 Thermopla
8	46	9.8	338100	1	TAC1D2	AL45064 Thermopla
9	45.6	9.7	14286	1	AE000747	AE000747 Aquifex a
10	45.4	9.6	11922	1	AE006657	AE006657 Sulfolobu
11	44.8	9.5	3382	1	DALIG	X63438 D.ambivalen
12	44.8	9.5	3382	6	A78674	A78674 Sequence 5
13	43.4	9.2	167000	1	AP000059	AP000059 Aeropyrum
14	42.8	9.1	9203	14	S51392	S51392 gag...rev [
15	42.8	9.1	10922	1	AE001529	AE001529 Helicobac
16	42.8	9.1	172552	2	AC080165	AC080165 Homo sapi
17	41.6	8.8	120920	2	AC010398	AC010398 Homo sapi
18	41.6	8.8	222400	2	AC009440	AC009440 Homo sapi
19	41	8.7	263335	1	CJ11168X5	AL139078 Campyloba
20	40.2	8.5	9202	14	S55323	S55323 gag...rev [
21	40.2	8.5	9202	14	VLWGAGA	L06906 Visna virus
22	39	8.3	9221	14	VLVIVIA	M60609 Visna virus
23	38.6	8.2	89047	6	AX067459	AX067459 Sequence
24	38	8.1	9225	14	VLVLVIB	M60610 Visna virus
25	37.4	7.9	9202	6	AL5114	AL5114 Visna virus
26	37.4	7.9	9202	14	VLWGA	M5543 Visna lenti
27	37.4	7.9	9203	14	VLWCG	M10608 Visna lenti
28	36.8	7.8	50487	2	AC068211	AC068211 Homo sapi
29	36.8	7.8	105288	10	AP003184	AP003184 Mus muscu
30	36.8	7.8	276523	10	AC012382	AC012382 Mus muscu
31	36.8	7.8	296950	1	AP001508	AP001508 Bacillus
32	36.2	7.7	37552	8	SC9302X	Z48179 S.cerevisia
33	36.2	7.7	163644	9	AC018377	AC018377 Homo sapi
34	36.2	7.7	164179	2	AC010910	AC010910 Homo sapi
35	36.2	7.7	175656	2	AC027761	AC027761 Homo sapi
36	36.2	7.7	282610	1	RPXX01	AJ235270 Rickettsi
37	36	7.6	36888	3	U29378	U29378 Caenorhabdi
38	35.6	7.6	147201	2	AC053469	AC053469 Homo sapi
39	35.6	7.6	174069	2	AC026852	AC026852 Homo sapi
40	35.6	7.6	177387	2	AC079680	AC079680 Mus muscu
41	35.6	7.6	180736	2	AC023925	AC023925 Homo sapi
42	35.6	7.6	225009	2	AC093366	AC093366 Mus muscu
43	35.4	7.5	160737	2	AL157816	AL157816 Homo sapi
44	35.4	7.5	189604	2	AC079459	AC079459 Homo sapi
45	35.2	7.5	163551	2	AC026520	AC026520 Homo sapi

ALIGNMENTS

RESULT 1
AP000007
LOCUS Pyrococcus horikoshii OT3 genomic DNA, 1485001-1738505 nt. position (777).
DEFINITION Pyrococcus horikoshii OT3 genomic DNA, 1485001-1738505 nt. position (777).
ACCESSION AP000007 AB009464 AB009521 AB009522 AB009523 AB009524
AB009525 AB009526 AB009527 AB009528 AB009529 AB009530 AB009531
VERSION AP000007.1 GI:3236134
KEYWORDS
SOURCE Pyrococcus horikoshii (strain:OT3) DNA.
ORGANISM Pyrococcus horikoshii
Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
Pyrococcus.
REFERENCE 1 (bases 1 to 253505)
AUTHORS Tanaka,T., Kawarabayasi,Y. and Kikuchi,H.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yutaka Kawarabayasi, National Institute of Technology and Evaluation, Biotechnology Center, 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail:genomeOT3@nitech.go.jp, Tel:++81-3-3481-8951,

REFERENCE
 AUTHORS
 Kavarabayasli, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, K., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Nakamura, Y., Robb, T.F., Horikoshi, K., Masuchi, Y., Shizuya, H., and Kikuchi, H.
 TITLE
 Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3
 JOURNAL
 DNA Res. 5 (2), 55-76 (1998)
 MEDLINE
 98344137
 COMMENT
 On Mar 17, 1999 this sequence version replaced gi:3131896
 gi:3131920 gi:3131974 gi:3131982 gi:3132032 gi:3132063 gi:3132118
 gi:3132139 gi:3132184 gi:3132233 gi:3133142
 gi:3132159 gi:3132184 gi:3132233 gi:3133142
 AB009464-AB009465, AB009521-AB009531: submitted (10-DEC-1997)
 Kavarabayasli, Y. is officially affiliated with the National
 Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
 305-0046, Japan.
 Robb, T. F. is at the Center of Marine Biotechnology, University of
 Maryland, Baltimore, MD, USA.
 Horikoshi, K. is at the Japan Marine Science and Technology Center,
 Yokosuka, Kanagawa 237-0061, Japan.
 Masuchi, Y. is at the University of Tokyo, Meguro, Tokyo 153-0041,
 Japan.
 Shizuya, H. is at the California Institute of Technology, Pasadena,
 CA, USA.
 The other authors are at the National Institute of Technology and
 Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
 All the sequence with length 100 codons or more between ATG or GTG
 and stop codon are defined as CDS
 Homology analysis is performed by Smith-Waterman algorithm against
 GenBank and GenPept release 103; EMBL release 52.0; SwissProt
 release 34.0; PIR-Protein release 54.0; and OML release 29.5.
 E-mail address for comments and questions: genome@nitech.go.jp
 Restriction map, ORF organization, sequence alignment and more
 information are available at w.w.w. site of Biotechnology Center,
 URL: <http://www.bio.nitech.go.jp/>.
 FEATURES
 source
 1. 253505
 /organism="Pyrococcus horikoshii"
 /strain="OT3"
 /db_xref="taxon:53953"
 /map="1485001-1738505 nt. position"
 complement(195..1253)
 /gene="PH1688"
 complement(195..1253)
 /gene="PH1688"
 /codon_start=1
 /transl_table=11
 /product="352aa long hypothetical protein"
 /protein_id="BAA30801.1"
 /db_xref="GI:3258117"
 /translation="MPPKVLIIAGNIGRATAMDLKDEVDVYIGVNNENLEKVEF
 ATPKVDASNFGLVEMKEFEELVIGALGFGFKSIRKAIKSKVDMVDVSEPNPL
 ELKDEAKQATVIVDAGAPGLSNILMGRIEQLDKEGIIYVGLPKDPKPLLYK
 ITWSPRLDEEYTRPARVIRNGKSVKVDPLSEKVKYKIGKFEFEAFISDGLKSMLETI
 NSRLEWTLRPHLEKIKVRLKGVLPENLDFTRVIEPLMRVETKDFSIMKVVG
 KGEGESEFFLDIEEDSMSSSRVGTFTAAIISRIVAENTCTFGVIPPEILGMREDT
 FRRIIDELKRGISIEGT"
 298..780
 /gene="PH1689"
 complement(3168..3392)
 /gene="PH1689"
 /codon_start=1
 /transl_table=11
 /product="160aa long hypothetical protein"
 /protein_id="BAA30801.1"
 /db_xref="GI:3258118"
 /translation="MQVFSATILEIIIAVNPVTLDEENMLSSSSYKRNKSISSPPFP
 TTFIIEKSLVSLVSKFSGLKPKPSLRLTIFSRCPGQSRVHSSLSSEFI
 VSNMLNPSLINASNFPIILFTLSESGSTFETFPFLITLAGLVYSIRSLGDHVL
 "

gene
 CDS
 871..1293
 /gene="PH1690"
 871..1293
 /gene="PH1690"
 /note="motif=prokaryotic membrane lipoprotein lipid
 attachment site"
 /codon_start=1
 /transl_table=11
 /product="140aa long hypothetical protein"
 /protein_id="BAA30802.1"
 /db_xref="GI:3258119"
 /translation="MLLRPCAKPASNTIVTWAFSSASSLSNGFSIGIKETSTMSTLDTFI
 AALMLNPNRPGKAPITNSNSFITSTSLSKLLASTFKGVANSFTFSRSLFTSPMYTS
 NSLSRSHATALPIEPAPKIRTFIGFTQIYFYACINFEY"
 complement(1297..1374)
 /gene="PH1691"
 /note="PH1691"
 /note="anticodon: GAC"
 /product="tRNA-Val"
 complement(1297..1374)
 /gene="PH1691"
 1464..2234
 /gene="PH1691"
 1464..2234
 /gene="PH1691"
 /codon_start=1
 /transl_table=11
 /product="256aa long hypothetical protein"
 /protein_id="BAA30803.1"
 /db_xref="GI:3258120"
 /translation="MTRSRKSFHILNVHMGKMGNOYEAIEKIKVIMRLIKKTYT
 YEDLSKITGLPTVLRNRYVRGKVLPSIERAKLEELKPYLNLEEEVKRRLKDFSGF
 FDTMSVLSTDTNLALIAEVALKYLRTGSKVLTAAATDGIPLAQVIANELGIDVYAK
 KRKEVGIEKFEYVNVPSAGSITLYLPMWAIKRGERVLIYDDVVRSGETORALVEL
 CHQADTLVGMFFLISVDVLDKLOEYINIPDALVKL"
 complement(2231..2686)
 /gene="PH1692"
 complement(2231..2686)
 /gene="PH1692"
 /note="similar to PIR:H64318 percent identity: 34.188 in
 117aa"
 /codon_start=1
 /transl_table=11
 /product="151aa long hypothetical transcriptional
 regulator"
 /protein_id="BAA30804.1"
 /db_xref="GI:3258121"
 /translation="MLKMLDLDLRKLEILLKDSSTVSREIAKDLNVAVGTYINRIK
 KLEDSGYICAFYTKLNTESIGYDLTALIGKAKCKKIREIERIAKDKRKYTCVIDYVG
 EDIIVIAKFRNEDMNRVYKGVLSIDGVEKTNHVVLVLEIKEDFRLEP"
 complement(2730..3155)
 /gene="PH1693"
 complement(2730..3155)
 /gene="PH1693"
 /codon_start=1
 /transl_table=11
 /product="141aa long hypothetical protein"
 /protein_id="BAA30805.1"
 /db_xref="GI:3258122"
 /translation="MEGELIAKLNKTGELEVIKEDFRFKCLDNCKGCCICNDIPLRE
 EDIERIKLGYDEYDFEVDTKMVRGPKFLGYTLKKRPFDNACVFLDPETKKRIYEH
 RPLACRLYPFALVRHGDLEIYVVRNVDVCPGHNPEGVTI"
 complement(3168..3392)
 /gene="PHS045"
 complement(3168..3392)
 /gene="PHS045"
 /note="similar to Swiss_Prot:P42179 percent
 identity:40.476 in 42aa"
 /codon_start=1
 /transl_table=11
 /product="74aa long hypothetical transcriptional
 regulator"
 /protein_id="BAA30806.1"
 /db_xref="GI:3258123"

```
/translation="MFILLVVRPGLENEVEYKLDKRPVEKEIYRLYGDYDIIRLSVE
GIKALDEPHDKVLRKLRGIEISETLIASSY"
complement(3398..4735)
/gene="PH1694"
CDS
complement(3398..4735)
/gene="PH1694"
/note="similar to PIR:E64312 percent identity:60.870 in
444aa; owl:AATFPH1 percent identity:50.236 in 430aa;
owl:SAPFH percent identity:50.611 in 415aa.
motif=ATP/GTP-binding site motif A (P-loop) ; SRP54-type
proteins GTP-binding domain signature"
/codon_start=1
/transl_table=11
/product="445aa long hypothetical signal recognition
particle protein"
/protein_id="BAA30807.1"
/db_xref="GI:3258124"
/translation="MVMVLDLSGRALSNALKKIARAGSVDEALVKEVVRDIQRALLQA
DVNVLVLKLETKERRALEKPPAGISKKEHIKIVYEELTKFLGTEAKPIEIKEP
TVLLTVQGSGKTTVAKLARYFOKRYKVVGVCSWTRPGAYHQLKQLDPYHIEV
FGDPNEKDAVLAKGEVYFTRVDLLIIVTAGHKEEKDLIEEMRMISSEIRPHEV
ILVIGTIGCOAQVNOALAFKATPIGSLIIVTKLSSAKGGALSAAVATGAPKEIGV
GEKIDDLPEFDFARFVSRLLGLDIOGLEKFEKEKEVEFTEEDIDRLRGKFTLKD
MYAOLEMRKGPQLQILRMIPGLGISLDELISVGEERLKRKFVIMDSMTEELMNP
EIIYNSRIKRIARGSGTSIKDVKELLTYQINOMKFFKSMNKRQLSLARRFCM"
complement(4767..5384)
/gene="PH1695"
CDS
complement(4767..5384)
/gene="PH1695"
/note="similar to PIR:F64336 percent identity: 42.778 in
196aa. motif=ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="205aa long hypothetical thymidylate kinase"
/protein_id="BAA30808.1"
/db_xref="GI:3258125"

Query Match      79.2%; Score 374; DB 1; Length 253505;
Best Local Similarity 87.2%; Pred. No. 1.5e-97;
Matches 410; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGACTGGAAATCAGAAAAGAAATACTTATAGAGCCATTTCTGAAGAA 60
Db 250855 ATGCTACTTCCGAGCTGGAAAATAAGAAAAGAAATACTTATAGAACCATTTCTGAAGAA 250914
QY 61 TCGCTCCACACGAGCTTATGACCTCAGCTGGCAGAGAGCGCTTTGTTAAGGGGAAA 120
Db 250915 TCACTTCAACACGAGCTTATGATCTAGAGTGGGTAGAGAGCTTTGTTTAGTGGAAAAG 250974
QY 121 TTAATCGAGCTGGAAGGAAGAAACGCTTATTCCTCCAGGAAGTAACGCTTAATC 180
Db 250975 TTAATGATGTGGAAGGAAGGAAGTAAGTATTCCTCCAGGAAGTAACGCTTAAT 251034
QY 181 CTAACCTTCAGAGGAGATAAAGTTGCCAGCATGTTATGGGGGATATGAAGTAAGGAGC 240
Db 251035 CTAACCTTCAGAGAGATAAAGTTACCTGATGATGTGATGGGAGATATGAAGTAAGGAGC 251094
QY 241 AGTTTACGAGAGAGAGGGGTTATGTTCTTTGCTTGGTTCACCCAGGATGGATGGA 300
Db 251095 AGCTTACGAGAGAGAGGGGATTTAGGTTCTTTTGGCTGGGTTGACCCAGGATGGGATGG 251154
QY 301 AACTTAACACTAATGCTCTACAAATGCCTCAATGAACCTGTCGAATTAAGATATGGAGAG 360
Db 251155 AACTTAACCTAATGCTCTACAAATGCCTCAATGAACCACTAGATTAATAATACGGAGAA 251214
QY 361 AGATTTGTCAGATCCGATTTATAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420
Db 251215 AGGTTTGTGACAGATGCTCTTTATAAGGCTAGAGGATCCACCAAGGAATCCCTATAGTGA 251274
QY 421 AACTATCAGGGGAGCACAGGTTAGCGTTTTCAGAGAGAAAGAACTCTA 470
Db 251275 AACTATCAGGGGAGCACAGATTTGGTATTTTCAAGAGAGAAAGAACTCTA 251324
```

```
RESULT 2
CNSPAX06
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source
gene
CDS
CDS
CDS
```

```
CNSPAX06 265118 bp DNA BCT 09-SEP-1999
Pyrococcus abyssi complete genome; segment 6/6.
AJ248288 AL096836
AJ248288.1 GI:5458960
complete genome.
Pyrococcus abyssi.
Pyrococcus abyssi.
Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;
Pyrococcus.
1 (bases 1 to 265118)
Heilig,R.
Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution
Unpublished
2 (bases 1 to 265118)
Direct Submission
Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Location/Qualifiers
1..265118
/organism="Pyrococcus abyssi"
/strain="Orsay"
/db_xref="taxon:29292"
164..1615
/gene="bgal"
164..1615
/gene="bgal"
/note="PAB2376"
/codon_start=1
/transl_table=11
/product="BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE)".
/protein_id="CAB50447.1"
/db_xref="GI:5458961"
/translation="MEFYGVVQSOFQFEMGDPYRRNLDTRSDMHVWRDPYNIKNEK
VSGDLPESGINNVELYEDHRLAKELGLNAYOLTIENSRIFPCPTYSVEVEVDGYG
FYKVKIRKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK
EYVNIKARAGWYDRAIIVEFSKFAAYVAMFKDKYVDFWATFDPMPVTAELGLAP
YVWPGLNPKAAKSLVMHQIIAHARAYDSIKFSDKPGVGLILIIIPALNPRDSK
HVKAENDLPHNRLFLVANKKLDIDINGEYVKVPHLKRNDWIGNNYTREVYKV
EPKYTEPLVTFVGVGEGYGSANPNSISPDNNPTDFGWEVYKGLYDSTAEAYEYSE
MYFTENGIAADSKDILRPYIVDHVKEVRLRENGINVLGYFHWALTDNYEWAMGFKI
RFLYEVDPIITKRIPIRRKRSVETIKVVKREGIE"
1834..2283
/note="PAB1019"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50448.1"
/db_xref="GI:5458962"
/translation="MPLPPDITFDSVTLLKMHSEKNKRLLEITLAKFTVYLSMLTVQR
YLTAAYLKRNVLELEFKDIYINIVPVSEETIILRASKTEANLIRKGVFLDLEDIITA
VTAETGSLIITDEPKKYEPIRRFGDITMPLDKFLREVELMVEKEII"
complement(2276..3079)
/note="PAB1341"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50449.1"
/db_xref="GI:5458963"
/translation="MFENLESFEAFKLLLEVEFLDNDPKQVKEKVLICIEPHPDCC
VIGMGTTIKLTERGIEVIYACMTDGYMGTLDSSLTGHELATIRREEESSKLLGVK
KIYWLNRDTLQYREYKDKLRIIRKPKDGVFLPDPLPYEAHPDHRTNGTGLALD
AVAFSPLPNFSNVDEIGLGHQVGFIALYTNKPNFYVDITDVMELKLAIRTHKSO
FPDDVWEVWEPLRTVALFYKKGKAGVKAEGFRIMPGLFYHITPADLI"
3192..4340
/note="PAB1020"
/codon_start=1
/transl_table=11
```

```
/product="hypothetical protein"
/protein_id="CAB50450.1"
/db_xref="GI:5458964"
/translation="MKVEVSVYKELVVLKGLTEDRIETLEMKGGIIIEDEFQIRYVR
FMSAGKLRGTVVDEEVVIFGPHIKRI INLRSGIRIRFKRGVEKPYGVNVR
VVMYKGLMIGTGGFICPTTFEIPQDFQPFQFNDPNLILVGMAGPEKPYLVNRP
PVKEDQFLPDVOEIKTGRSLPVEERLKAIEVGINHVEVFGKTKDDVDLVLOLI
ERLSEGRGIIIMKSPDMKKIVKVTPTYANINDIKIGARVFEVLPQVETSRISLAF
VLAERKIGFEFERVAKELGSAILOPVESEIDVQEOEDIEHLFKVRKRIETAYKMV
THFEKLGKIEIVDIEEIKDGRWRIIFKRLYPDATNEIREILGGKAFVD"
CDS
4351..5658
/note="PAB1021"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50451.1"
/db_xref="GI:5458965"
/translation="MEDILRLVKAIRAADPKYKAVINSRVVEGKLLVKDEFRIGK
KYVLAFGACQMDARAIEILDVEGVATYKYGKGLKKRYIEAGHPIDENSIR
GATEGLKLEKVEDIVFLISGGSSALFELPDGIGLSLEDKRLNELLKSGAKIHE
INTVRKISKYGGLAKKIGIGIVLLISDVGDNLGAISGPTVKDPTTFQDAKRI
LELIGWDRPESVRSHIEKGINGEAETLKEELPNHNFLLIASNISQCAVVEEATK
LGRAYVMTTLEGAKDGLFISIVQIEAKTGNPFEPVVLVFGGETTVYIEGEGG
KGPQGLSATSARKIKGNALIVAFDPTDGTGPTDAAGGIVDGETYEKLRRKGIDIE
KVLKHNSTEALKKVGSLTGTGPTGNVNSMIIAIIITSKQGR"
complement(5633..6655)
/note="PAB1338"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50452.1"
/db_xref="GI:5458966"
/translation="MPIVEARVKVGISSWVTSKVSWRDALAOLESDRVIVKYLRMGE
VTGDSFPFSALTDIGIRPDELKNPEKDFGLFYVPGGELIVFTIEENLLIYD
EKKEAFVHKVEVLINGTKVLMQLAIRVGGAIKNMSEWEGWLRVIVKYSARTOYE
RSVVITQDKRVSIFSOLEIDEIEVEMNGKKVRAWKIRHFKHINOSVTSYLYIPDKK
VOLFVRLYLLKTPSAMEFI IKIADDFPTLKSEFOELMEKLESELDMEKQILVA
LYSGINPLEHOLFGLTTEKEIEIYDVRMDKGLLKVIRKIVDLTNEGRKIVNLLK
YGLVSN"
complement(6670..7425)
/note="PAB1337"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50453.1"
/db_xref="GI:5458967"
/translation="MDVGTILIAVVLIGISTVSAWIVSSVISQOMKMOALMISRL
EVTETVEKKRGTGRGSEAKPEVAKQEVDRVSEDIYVKIKQIIDEKVHNVANRKL
KRELMLLDVAGVAGLGVSEVNAFLMRVSELDEPKRLMLAAYPEPKKRLNQ
MTHVARTKLPTVEKKGKGVSKLSPDEALIRITSSIKSAISILDDLIKRGDNP
AVTPLEIKLSQLEKCKKVEDEKRLQLEAL"
complement(7468..9687)
/note="PAB1336"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein (tlpC)"
/protein_id="CAB50454.1"
/db_xref="GI:5458968"
/translation="MQFRKKIILMAIMSAVITMAIPGGVTIYFSFNKVIYVIGETOAP
/LEQGGQLAMSAEAEKMFDDFARVADYGTMAEVKAEYAKGLDNETAMREFL
SEFKIKELNKDVSIVYVEDEGHMYPDPDEPGYDPVRPVPVYQAKEGGAIITE
PPDKATQGWITVYVYVNGQFVGILGVFVSTLLNETKKIKLGTGTIADIDQ
GMTLVHPNEEFINKINIVPELKDADLKKKDEGVLYTFQGIKKIGYKRLKTT
GIVVYPLSDITGPILSATDIAINKMKMVIQGFVGVFIIVAILVVIYKASNTL
AFLEKLKAAQALAEGRILKQVSEYLUKIRYLERDEITGALIOAFEAIVSKDLVGLTNAIS
KKLERLAEGLSNGTVEVGRGLRDI IEDIRVETTFRESIGSLVEMANDLEKRNAL
AOVSDVTEANEGSQKQDALKRIEDIQHMMSRIEETVSKVAEMSRIEITNVTISIA
EQTNLLALNAIETAAEAGAGRAVVAQIEKRLAEESQQAADNKISIDITDEIKE
RAVENLAASAEETTASAEVSAIEQQTAAIEELRRAAQLKDMVGRMQIVGKPKV"
complement(9704..10195)
/note="PAB1335"
/codon_start=1
/transl_table=11
/product="protein-glutamate methyltransferase (cheB)"
Query Match 59.6%; Score 280.8; DB 1; Length 265118;
Best Local Similarity 75.0%; Pred. No. 1.7e-70;
Matches 351; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
QY 1 ATGCTACTTCCGACTGGAAAATCAGAAAATAACTTATAGAGCCATTTCTGAAGAA 60
DB 262045 ATGCTCTTCCGACTGGAAAATAAGGAAGAGATTTTATAAGCCATTTCTCAGAGAA 262104
QY 61 TCGCTCAACACGACAGTGTATACCTCAGATGGGCGAGAGAGCTTTTGTTAAGGGGAAA 120
DB 262105 TCGCTTCAACGACTGGCTACGACTTAAGGTAGGCAAGGAGCTTACATTCAGGAAAG 262164
QY 121 TTAATCGAGCTGAAAAGGAAGAAAAGTCGTTATTTCCTCAAGGGAATACGCTTAACT 180
DB 262165 TTICATAGATGTTGAAAAGGAGGCAAGTCATAATACCTCCAAAGAAATATGCCTTAATA 262224
```


QY 172 GCCTTAATCTACCTCGAGAGATAAAGTTGCCCCGACGATGTTATGGGGGATATGAAG 231
 Db 359 GCCTTAGCCATCAGTAGAGATATTTAAATGCCATAAGACACTTTAGCGATTGTTA 418
 QY 232 ATAAGGACAGTTAGCAAGAGAGGGTTATTGGTCTTTTCCTTGGTGGTTGACCCAGGA 291
 Db 419 GGGAAAGACACTTACGCTAGCTAGGTGGGATATTGTGAATGTTACGCTTTTGAGCCGGAA 478
 QY 292 TGGGATGGAACATTACACACTAATGCTCTACATAAGCTCCCTCAATGAACCTGTCGCAATTAGA 351
 Db 479 TTTGAAGGCTATATTACGATTGAAATTTCTAACACACACTAATTACCGGCTAAAGTCTAT 538
 QY 352 TATGGAGAGAGATTGTGCAGATCGCATTTATA 384
 Db 539 GCCAATGAGGGATCGGCAAGTGGTGTTTTA 571
 RESULT 7
 APO00993 239340 bp DNA BCT 07-JUN-2001
 LOCUS Thermoplasma volcanium genomic DNA, complete sequence, section 3/6.
 DEFINITION APO00993 BA000011
 ACCESSION APO00993.2 GI:14324830
 VERSION
 KEYWORDS Thermoplasma volcanium (strain:GSSL) DNA.
 SOURCE Thermoplasma volcanium
 ORGANISM Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae; Thermoplasma.
 REFERENCE 1 (sites)
 AUTHORS Kawashima, T., Yamamoto, Y., Aramaki, H., Nunoshi, T., Kawamoto, T., Watanabe, K., Yamazaki, M., Kanehori, K., Anano, N., Ohya, Y., Makino, K. and Suzuki, M.
 TITLE Determination of the complete genomic DNA sequence of Thermoplasma volcanium GSSL
 JOURNAL Proc. Jpn. Acad. 75, 213-218 (1999)
 REFERENCE 2 (sites)
 AUTHORS Kawashima, T., Anano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshi, T., Yamamoto, Y., Aramaki, H., Makino, K. and Suzuki, M.
 TITLE Archaeal adaptation to higher temperatures revealed by genomic sequence of Thermoplasma volcanium
 JOURNAL Proceedings of the National Academy of Sciences of the United States of America. 97 (26), 14257-14262 (2000)
 MEDLINE 20570466
 PUBMED 11121031
 REFERENCE 3 (bases 1 to 239340)
 AUTHORS Suzuki, M.
 TITLE Direct Submission
 JOURNAL Submitted (20-DEC-1999) Masashi Suzuki, National Institute of Bioscience and Human Technology, Core Research for Evolutional Science and Technology, Centre of Structural Biology, Higashi 1-1, Tsukuba, Ibaraki 305-0046, Japan (E-mail: nanano@nslh.go.jp, URL: http://www.aist.go.jp/RIODB/archaic/, Tel: 81-298-61-6582, Fax: 81-298-61-6041)
 COMMENT On Jun 6, 2001 this sequence version replaced gi:11990452.
 The most recent annotation is available at
 http://www.aist.go.jp/RIODB/archaic/.
 FEATURES
 source Location/Qualifiers
 1. 239340
 /organism="Thermoplasma volcanium"
 /strain="GSSL"
 /db_xref="taxon:50339"
 1. 3609
 /gene="TVG0609019"
 1. 3609
 /gene="TVG0609019"
 /note="unknown product"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /protein_id="BAB59757.1"
 /db_xref="GI:14324831"
 gene
 CDS
 1. 3609
 /gene="TVG0609019"
 /note="translation elongation factor EF-1 beta"
 /protein_id="BAB59760.1"
 /db_xref="GI:14324834"
 /translation="MIEMADVLVSFKLLPSDSDDVTSMSEVKEKLVGCKINIEE
 KDIGFLKYIHLEIVEDGEVDRIEKVLMSVKGVGEINTENVSII"
 5924. 7267
 /gene="TVG0612677"
 5924. 7267
 /gene="TVG0612677"
 /note="unknown product"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /product="phosphomannomutase"
 /protein_id="BAB59761.1"

/translation="MVDKALTKNLMIKINAIITAEKDQDKKQRIIDDLISLLRNTNPPD
 IDKDEIPVSVLRNVSDENQRELITLDSIVPQNGKLETKVVDLTKNGYIDVV
 KTIITSDNYFCSIDSIISLLVSSGHIERDAAALYSLKCRYEERIKIQVMYARNRPD
 KTAITEGVNNGNEMAIKFTIMEINSRQSELYLGPLNCSIKTDDKQGMRAALDFA
 LKIDISADALIKLQVAAALSGEYTTIGKELAKNGKLKPADKLMKALCLSKSGNY
 GAIETYSKDLTSPDIDATIDLMDLLEYHGKASELYLVVEKIGRNQLRDLYIRLI
 QLMKAGSDIGKAVELAKAYKKFPSEDDVIEIYARIVDEIGKNGSFDAKSLDKGG
 KEAQSFILNLFSEREYEFVKRPFESDDVIEIYARIVDEIGKNGSFDAKSLDKGG
 ENKLLDDPDVTDSVYKVRKDEALEILNSLDIKGYSVSVLDRLQGRIGRSDYLLD
 YAVSKCSKIATYSVYFPRMISLDPKVSALSTNCLSEIYDFILTKIKGTNSG
 IYEDPRYLFCFSFTIDAGKLDLADYKLLQREKVEYDFYFTYAKLLKSGRNRKAS
 RVKAELELFPNIDFHLHLNDDPESLESIOAIFDIDAMIYDVYQSLIEDHPOLV
 SEAVNFLISRGIDDDVRLRLSDRVYLNKIDIGALKVSSRIISOSSESDIAIHYRIL
 KEIDLAALAFMKNNDITENGKEELIEAGELMYSKRYEDAISAFKALRGVNPISI
 PHYLSLIETGSAEADLVRQISMDVYAVKLPFREGKISELVNLTQTOEKDNLV
 RYIVQGGWNPDKKATFDYKSTGDTGTEIGKIAKRLNENKLEALELRIQHDDE
 SPAEASIMTDLRRQGLTEAVEIAAASAOCSDECEKLEYRRTYSIMYDLGMYSO
 TVIKPKKRAEKDNNVTPYVVKSMVELKQFDEAEVLSFSTLPEAVKNALESYINOG
 YDLRNLKYASRVLSHEFKAGKIMPIDESVKKSDVPEMDIQRDFTLTKYIKGVNEN
 KLEELSGDVIRKIALKLIKLSLSSIKINFYAFLODDSAIMAKNLVLYIKRAKYTHI
 PTOSDVEARISFSFSEATSNPIDVIKRYGVGKIKKALIEAAKSLQNEGSLKON"
 3654. 4232
 /gene="TVG0609642"
 3654. 4232
 /gene="TVG0609642"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /product="hypothetical protein"
 /protein_id="BAB59758.1"
 /db_xref="GI:14324832"
 /translation="MYGNLNNEEKDILVSLALTAAGIAFRFRSPYGPFIPLAMAF
 LYSVTAFLMHELAHRYVARSYGIAYFKMPLGLLALITVFGFIAPAGVNIIGGL
 YKQDQIGKTAGLAPAMNIFGLIFGLSHITVPVVALIFGYVAAAMNFWFGFNLPLI
 PPLDGYKVMNWDIYVFLVSIILALIFVALVF"
 complement(4263. 5321)
 /gene="TVG0609673"
 complement(4263. 5321)
 /gene="TVG0609673"
 /note="unknown product"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /protein_id="BAB59759.1"
 /db_xref="GI:14324833"
 /translation="MVVYVSIASLIAGVFSYIGIYKSNELKORKNCRPKTLVTPCKG
 LDGLLEDNLKAITKONYDRADYLVGVVDSLODESVSLEKVGINYIISDYTCBCKSGKV
 KAIATAIKYTDYDAVIAADSDIRPNSNMLEDLVSLCNDRVGLSTPFPYFPIGGLA
 SKVAVMCTVGGSLMESKLTREGWGSGLAFKSLISDKMDYETSLSDSALTAICKS
 KLDIATYKAEAPAIYTAENFQFQFWANQFALISINASPRVETVGIIEYSLYIEFF
 SAVLTSISLAFVYLFIPVIGVYRVYERSVKSPIAFIYLIFFIPIYLANLLKAKRM
 GEIINWGRQIKLQKLDLP
 complement(5330. 5605)
 /gene="TVG0610740. 5605"
 complement(5330. 5605)
 /gene="TVG0610740"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /product="translation elongation factor EF-1 beta"
 /protein_id="BAB59760.1"
 /db_xref="GI:14324834"
 /translation="MIEMADVLVSFKLLPSDSDDVTSMSEVKEKLVGCKINIEE
 KDIGFLKYIHLEIVEDGEVDRIEKVLMSVKGVGEINTENVSII"
 5924. 7267
 /gene="TVG0612677"
 5924. 7267
 /gene="TVG0612677"
 /note="unknown product"
 /codon_start=1
 /transl_table=11
 /evidence="not_experimental"
 /product="phosphomannomutase"
 /protein_id="BAB59761.1"

```
/db_xref="GI:14324835"
/translation="WAVRLFGTNGIRGTPNDMSPDFAMSGVKGATGYVSGEIAFGT
DPLRGDMIAKNVAGILSGISVLDGLVLPALQYXSKHGIIYAVVITASHNPOY
NGIKVIDKDGTELRGPKGEIEIYFSKAFRVVKNENIGRIQVSDAIDIYINGVISN
VDSNLIKSHAFVLDAGNGAAYSTPPKLEKLGCSITTLNANPDGRFTARNAEPVE
NLVNIISLMKSRFLDGAHDGADRAVEIDKGNFIDGRMLALITKYIAKAGDVVV
TPVSSDTICNAIGAKRIETKYGAPIVARAMINDGATIGGEENGVIYKGHQYCR
DCAMTVSVILNMASTNKLSELISLPVYLIKRVRRVDPWSLARELKKGERVDE
TPGIKLDGWDGILRPSCTEKILRIYAHSSDFELANKYCEQYIRKVNIDIQKLS"
gene
/length=7259.7858
/complement(7259.7858)
/genes="TVG0612669"
/feature(7259.7858)
/feature(7259.7858)
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="phosphatidyltransferase"
/protein_id="BAB59762.1"
/db_xref="GI:14324836"
/translation="MLDSYRAKVDSILTPIAQALGFLSPNTISIIISLAGLTGIFY
LNVYLSLFLIISALFADVCKLARIQKSKRGDFIDHSFDRFSDVFLILGTFESR
YANVDLGFALLIGVILTSVMGTQQAALGLNRLYAGIMGRAIDLVLVPIALQIPIGI
IVRFELITPNTMLLIRYGIUGSATVQVRFIIFKELR"
complement(7878.8339)
/genes="TVG0613288"
/complement(7878.8339)
/genes="TVG0613288"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB59763.1"
/db_xref="GI:14324837"
/translation="MGKIACIYCPGAGKSTVCSKLRGVYCNCKEKNELAKYGCFLFD
EVDVCELEKLAERDFGICSHYSLGSCSTVFLEADLNDLIDMRARGYSEKI
QNIETQMSIYFYSLESLRIPANRIFTLYNGNIDETAKRIIISRSRNNK"
complement(8378.9715)
/genes="TVG0613788"
/complement(8378.9715)
/genes="TVG0613788"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="cytosine permease"
/protein_id="BAB59764.1"
/db_xref="GI:14324838"
/translation="WTEHKFGDIETKRIIKGALSDRIQNPSSMYIWFASNLTVG
DFALGPIYGLSESIALSVLGLTAGGILLAYMSLGSINHPQMDMGKPGFRI
GGSFALLQWNTAGVILNVNLSFALYEVILKIPYFIILLIIVALTALGVRAI
RALERYMSVGLIFVILFVWLRYHTLIYSGSSVSPIPAFGITFASAFYIMSWG
PVASDYRHOYSPKSRKVFLLTGLSALATVFAETVGMFVAGISGPNPSTADLEL
VMSFALAGMISFLGGTSANAINLYSNSLSFIANGFKTRRTIGVILATFLSVILGIF
GTRYGYFETFLVDLYWITPWIGVMTAHFFLVHPNTSTMNTKLNMAGVLSYLIISI
LISIPFMSFAGLFNPIIASILGGVDISYFVSFTSMILYVLYGLKLENKLAIV"
9787.10302
/genes="TVG0615712"
/9787.10302
/genes="TVG0615712"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB59765.1"
/db_xref="GI:14324839"
/translation="WHIRYDQAMDDETKVEAILYATRNPLSVRSISLILGIEAGAIRS
IILKRLRYKRNSTLEIAKIGNKYRIOLKKEYDYFAYRVMEPELSKYTGFLATVAL
NGASLSFRKRYGSRDMDISKLTMSLIRTSKKGNGTAIYIGENFVKVFGITKKEL
LERATEYDRQN"
10411.11718
/genes="TVG0617128"
/10411.11718
/genes="TVG0617128"
/codon_start=1
```

```
Query Match 10.1%; Score 47.4; DB 1; Length 239340;
Best Local Similarity 50.2%; Pred. No. 0.0093;
Matches 117; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 94 GGCAGAGAGCGCTTTTGTGTAAGGGGAAATTAATCGACGTGGAAGGAGAAAGTCGTT 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22761 GGTACGATGTTAGAGTAGAATCTATACTTGCAGACGAGAAGGAGTTGATGAACACTATT 22820

QY 154 ATTCTCCCAAGGGAATACGCTTAATCCTAACCTCGAGAGGATAAAGTTGCCCGACGAT 213
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22821 GTTGGGACGATGAACACTTTTCTTGTCACACCTTGAAGATATTGAGGCTCCCGAGAAT 22880

QY 214 GTTATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAGAAGGGTTATTGTTCTCTTTT 273
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22881 GTCATGCCCAATATTGGATTCTGTTCAAGTTACGCCAGGCGTGGAGTAATAGCTCTCTTT 22940

QY 274 GCTTGGGTTGACCCAGGATGGGATGGGAACTTAACACTAATGCTCTACAATGC 326
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22941 GGCACAATAGACCGCGCTATCATGCGAGCTTACATTATATCTTTTAAATGC 22993

RESULT 8
LOCUS 338100 bp DNA BCT 04-OCT-2000
DEFINITION Thermoplasma acidophilum complete genome; segment 2/5.
ACCESSION AL445064 AL139299
VERSION AL445064.1 GI:10639496
KEYWORDS
SOURCE Thermoplasma acidophilum.
ORGANISM Thermoplasma acidophilum
Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
Thermoplasma
1 (bases 1 to 338100)
Ruepp,A., Graml,W., Santos-Martinez,M.L., Koretke,K.K., Volker,C.,
Mewes,H.W., Frishman,D., Stocker,S., Lupas,A.N. and Baumeister,W.
The genome sequence of the thermoacidophilic scavenger Thermoplasma
acidophilum
Nature 407 (6803), 508-513 (2000)
20479972
2 (bases 1 to 338100)
Ruepp,A., Graml,W., Santos-Martinez,M.L., Koretke,K.K., Volker,C.,
Mewes,H.W., Frishman,D., Stocker,S., Lupas,A.N. and Baumeister,W.
Direct Submission
Submitted (28-SEP-2000) Max-Planck-Institut fuer Biochemie, Am
Klopferspitz 18a, D-82152 Martinsried, FRG Bioinformatics, Smith
Kline Beecham Pharmaceuticals, Collegeville, Pennsylvania 19426,
USA GSF-Forschungszentrum fuer Umwelt und Gesundheit, Munich
Information Centre for Protein Sequences (MIPS) am
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152
Martinsried, FRG Project Coordination: Andreas Ruepp and Wolfgang
Baumeister E-mail: ruepp@biochem.mpg.de, baumeister@biochem.mpg.de
Information on performance of analysis and a more detailed
annotation can be viewed at:
http://www.biochem.mpg.de/baumeister/genome/ and at:
http://pedant.mips.biochem.mpg.de/.

FEATURES
Location/Qualifiers
1..338100
/organism="Thermoplasma acidophilum"
/strain="DSM1728"
/db_xref="taxon:2303"
115..1704
/genes="Ta0324"
/genes="Ta0324"
/genes="Ta0324"
/notes="similarity to known protein: O-sialoglycoprotein
endopeptidase (EC 3.4.24.57) [validated] - Pasteurella
haemolytica (serotype A1); PIR2:A38108"
/codon_start=1
/transl_table=11
/product="O-sialoglycoprotein endopeptidase related
protein"
/protein_id="CAC11469.1"
```


LTATATATSELIEDVNHVNHVKNVQVRNQLDLASAKIAPGVNVRVYVNTGDKIDKNILY
LTVTSTAEHGDGTMGRNKGVLITPMRPMSEATAGKNPVNHYKLYNVLNLANLIA
NKIAQEVKDVXFSQOVVLQIGRPIIDPLIANVDVITYDGLTDTETKNEISGIVDEML
S5FNKLTLELLEGGATLF"

/gene="snRNP-1" .9866)

/note="SSO5410"

/complement(9603. .9866)

/gene="snRNP-1"

/note="Transcription, RNA modification"

/codon_start=1

Query Match 9.6%; Score 45.4; DB 1; Length 11922;
Best Local Similarity 48.1%; Pred. No. 0.022;
Matches 161; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 121 TTAATCGAGCTGAAAAGGAAAGTTCGTTATTCCTCCAAAGGGAATACGCCTTAATC 180

DB 471 TTCTTCCAAACGGAGAGAGAGAGTATTAATCAACCTTATGAGCATGTACTACTA 530

QY 181 CTAACCTCGAGAGGATAAAGTTCGCCGACGATGTTATGGGGATATGAAGATAAGGAGC 240

DB 531 ACAGCTGAAGAATATATAGATTAAATATGATGAATGCTTCGTAATCTAAGGTCA 590

QY 241 AGTTTAGCAAGAGAGGCTTATGCTTCTTCTGCTGGTTCACCCAGGATGGATGCA 300

DB 591 ACATTGCTAGATTAGTTTATTCACCTACATAGTCTGATCTGGATTAAGGA 650

QY 301 AACTTAACACTAATGCTCTACATGCTCAATGAACCTGTGCAATTAGATATGGAGAG 360

DB 651 CAAGTAACTAAGAGTAGTG---GGCTCATCTTCCCTGTAATTAAGAGAGACACA 707

QY 361 AGATTGTCAGATCGCATTTATAGGCTAGAGGTCGCGGCAAGAACCTTTACAGAGGA 420

DB 708 CGTTTATTCACCTGATCTTTCCTCCAGACGCTACACACAGTTGAATATCCGTACCAAGGA 767

QY 421 AACTATCAGGGGAGCACAGGTTAGCGTTTTCAAA 455

DB 768 AAATATCAAGGGCAAAAGGCGTTACATTACCCAA 802

RESULT 11

DALIG/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

gene

CDS

DALIG 3382 bp DNA BCT 29-JUL-1993

D.ambivalens lig gene for DNA ligase.

X63438 S47845 S47846

X63438.1 GI:40784

DNA ligase; lig gene; thermoacidophile.

Acidianus ambivalens.

Acidianus ambivalens

Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Acidianus.

Kletzin, A.

Direct Submission

Submitted (06-DEC-1991) A. Kletzin, Max-Planck-Institut f

Biochemie, Am Klopferspitz 18a, D-8033 Martinsried, FRG

(bases 1 to 3382)

Kletzin, A.

Molecular characterization of a DNA ligase gene of the extremely

thermophilic archaeon Desulfurolobus ambivalens shows close

phylogenetic relationship to eukaryotic ligases

Nucleic Acids Res. 20 (20), 5389-5396 (1992)

93065206

Location/Qualifiers

1..3382

/organism="Acidianus ambivalens"

/strain="DSM 3772"

/db_xref="taxon:2283"

complement(250..771)

/gene="ORF3"

complement(250..771)

/gene="ORF3"

/codon_start=1

/transl_table=11

/protein_id="CAA45033.1"

/db_xref="GI:40785"

/translation="MILGDRDLKYYLEKRGWIVISPLTQDTIRENGVDLVRGVEIARFK

KTDEIYEDGKDPFRSFYEIEKGEDEFIYNEHVLVLYEYVKLPNDVMAFVLRSSFA

LGLFVPTIVDAGFEGQLTIEVLGSAFPVKIKRGTRFLHLIFARTLTVPENPYHGKYQ

QQQGVTLPKFKFR"

787..883

/note="lig box A"

/note="lig box B"

/evidence=experimental

908..2710

/gene="lig"

908..2710

/gene="lig"

/codon_start=1

/transl_table=11

/evidence=experimental

/product="DNA ligase"

/protein_id="CAA45034.1"

/db_xref="GI:40786"

/db_xref="SWISS-PROT:O02093"

/translation="MEFKIAYEDRLEKISSRLQTSLSLADLEKKTQKNVLDKVVYI

IQGLWDFLGNPELGEKELIIRALSSTATSYSDDEIEKNKSVGLQGVAFIKQKQ

QASLAFLEAGAKSKPLTVEKYDDLAATFTGGGONFPIVERAYNLRADLGNLAKILANG

KILVFDGRLRVGIGDRIIDALATFPGGONFPIVERAYNLRADLGNLAKILANG

GIEQKNIKPQGPPIRPMLEKSLKSVNIALVDYKYDGERGQIHKAGDKI

FISKRLENTNQYPDVAEYISKYVKNFEIIEVEIEIPDPEPTEMRPFQELMHRKRI

SDIHEATKEYPNVFLFDLMYIEGEDYTVKPLSERRKKLESIVEDNDYVHATHITLD

NVEKLKFFQAISEGAAGVMVSLAPDAIYOAGSRGWLWIKFKRDYSEMAVTVDLV

MVGAFLHGRGKGYSSFLMAAYNPDPFVETVCVKSAGSTDAELDDLOKIAELKRD

TPHPRVSTMPVDVMTLPALVAEIIIGAEITISPLHTCCCKDQYAEGLLSIRPFRTRWR

PKSPEDATTNREILEMYKSQLKKEEKPSQSV"

2961..3382

/gene="ORF2"

2961..3382

/gene="ORF2"

/codon_start=1

/transl_table=11

/protein_id="CAA45035.1"

/db_xref="GI:40787"

/translation="MGISYGVKIAIEDEKIELVQADHVFVGAQVVTNNGESVGYTG

DFKNPGKTPILNTDVLIVDTYCKPTFRFRKFRDEVEVLFSVDYVNDLSLIYGPVRVAY

YKIQEAMKILKNGISAPFIVDGKIKIEVTDAIKYGL"

BASE COUNT 1188 a 512 c 683 g 999 t

ORIGIN

Query Match 9.5%; Score 44.8; DB 1; Length 3382;
Best Local Similarity 47.9%; Pred. No. 0.028;
Matches 161; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 120 ATTATCGACGTGGAAGGAAAGTTCGTTATTCCTCCAAAGGGAATACGCCTTAAT 179

DB 598 ATTTTATGAGATAGAGAAGGAGATGATTCATCATCTTACCTAACGACATGTACTGTT 539

QY 180 CCTAACCTCGAGAGGATAAAGTTCGCCGACGATGTTATGGGGGATATGAAGATAGAG 239

DB 538 AGTTACTGAGGAGTATGTTAAACTCCCAACGACGATGCGCTTCGTCAACTTAAGGTC 479

QY 240 CAGTTTAGCAAGAGAGGGGTTATGGTTCTTTTCTGGGTGACCCAGGATGGATGG 299

DB 478 TTCTTTTCCAGGCTGGGTCTTTTGTACCCAGGCTACTCGTAGATGCGGCTTTGAAGG 419

QY 300 AAACCTTAACACTAATGCTCTACAAATGCCTCAAAATGAACCTGTGCAATTAAGATAGGAGA 359

DB 418 ACAACTCAATAGAGTACTAGGTTCTGCTTTT---CCCGTGAAGATAAGAGGGGGAC 362

QY 360 GAGATTGTCAGATCGCATTTTATAAGGCTAGAGGGTCCGCAAGAACCCCTTACAGAGG 419

[illegible]

```

CDS
complement(1152..1490)
/gene="APE0276"
/notes="similar to PIR:F69324 percent identity:47.115 in
104aa."
/codon_start=1
/transl_table=11
/product="112aa long hypothetical protein"
/protein_id="BAA79229.1"
/db_xref="GI:5103913"
/transl_table="MRAGFIRVEEVDRRLVSLRFEFVDTGAEAAIVLAFERNAGLLL
MDDRARNLAKLGLQVMGTGLGIALAKYKGLTSKAKPIIDKLIESGFWSRRTLEEF
LRKLGESRNP"
1228..1722
/gene="APE0277"
1228..1722
/gene="APE0277"
/codon_start=1
/transl_table=11
/product="164aa long hypothetical protein"
/protein_id="BAA79230.1"
/db_xref="GI:5103914"
/transl_table="MSLSIIIGFALEVNPLYLANAITSVPITCNPSFLARFLARSSSI
SKPALRSKARTIASASPVSTNSKRDTSLSTTSSTRIKIPALTFTSLPGLPATVISA
QTVSGTRSPKSFRRSRHRTLASAMIMLETTSRSSSSLLLCKLIGLQILQLL
LAIV"
complement(2280..2756)
/gene="APE0278"
complement(2280..2756)
/gene="APE0278"
/codon_start=1
/transl_table=11
/product="158aa long hypothetical protein"
/protein_id="BAA79231.1"
/db_xref="GI:5103915"
/transl_table="MDKDEKDGALVLDITNIILASVLRHSGSYTRQVVIYLDVLGK
FTSEKALHDEHLEELKRYKVSVEELRAAIRLLLVNVMVEKGTYNAYMEKAGEC
VKDLADADFALALHLSKYNKVIILLTNKNKYIENCLKAYGVILYTPSDLRVPVN"
complement(3301..3669)
/gene="APE0279"
complement(3301..3669)
/gene="APE0279"
/notes="similar to PIR:H69385 percent identity:39.175 in
97aa."
/codon_start=1
/transl_table=11
/product="122aa long hypothetical protein"
/protein_id="BAA79232.1"
/db_xref="GI:5103916"
/transl_table="MVGEIERSISVYAPRLFLVEVAGVLYRYLAPSIVERVLDAFSS
KVIILGDEAYRIAEVAYALAGSRGADSYILGLAKTLNLPVATSDKVQAQNAKAGIK
SFYLSNNLEELKRYMECK"
complement(3752..3979)
/gene="APES010"
complement(3752..3979)
/gene="APES010"
/notes="similar to PIR:F69385 percent identity:38.462 in
52aa."
/codon_start=1
/transl_table=11
/product="75aa long hypothetical protein"
/protein_id="BAA79233.1"
/db_xref="GI:5103917"
/transl_table="MELRRVEIGMSKVIRVRYRGVLKPLPVPVNLEDGEEVDIIIREN
LAELARRIRRLSOREEPEILSRERSLA"
4028..4402
/gene="APE0280"
4028..4402
/gene="APE0280"
/codon_start=1
/transl_table=11
/product="124aa long hypothetical protein"
/protein_id="BAA79234.1"
/db_xref="GI:5103918"

gene
/transl_table=11
/product="108aa long hypothetical protein"
/protein_id="BAA79235.1"
/db_xref="GI:5103919"
/transl_table="MVHTPTVTLSPPMTSSVGVWRWAGTVPAGNATEHTAVPAISO
SSDFSTFCASRTSIAASSAVVSALPTPTPATILTSSTKLILGLYFSPSHDKIRKAK
KROKAP"
complement(4803..5465)
/gene="APE0282"
complement(4803..5465)
/gene="APE0282"
/codon_start=1
/transl_table=11
/product="220aa long hypothetical protein"
/protein_id="BAA79236.1"
/db_xref="GI:5103920"
/transl_table="MSSGILFQYRSNICVYFFHFGSSLFASLIISIMIDIKGFLIW
SAYARVLPWYFLAGACAI SFAYSQLAFRMWYILSLFPCFRASRSIAVSIQGG
LISLFLNSWYSFMFCMISFLMGSSASFLRSRSRYALSACSCSVASSIVSTMY
WSICRSVSGGALSSMCCPFFLRLSRSMSPAPRPISSLLILPLWSPIQLRUCY
PR"
4831..6840
/gene="APE0283"
4831..6840
/gene="APE0283"
/notes="motif=ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="669aa long hypothetical protein"
/protein_id="BAA79237.1"
/db_xref="GI:5103921"
/transl_table="MGDRGDRMSRELIGIGAGILITLLLNKLNKKOHLERAPPO
LTREIDQYIVDTTIDEAYEOLAEYDPAVELLKRSKEADPHIKLIONIKYVHEL
FKDNERMPKMLTALILEALKHGVKRODKIOYILKASHEYAKETAQALPAKYOG
RYRAYADOMRNPFIILFAMIREANRODDPKKYTOTMFDLYNKIPDDIYRAOS
OHEVTYKOLLLEALREAESGVFPHKTREILIGICGEGPAGAGELIAYDOR
YPTLAOKVILKTLVGLFELGEGGVHEDLALNLDSPITTHNYFFLAEMFENLE
KGRHVSPILOLLOLPVNPDELQRAYRIQIEQCKSISEIPEERKAFREVIERLAN
EIMNRPFFVIEELTYATGMNIDAKLILIGVLDLDFAKIYSGEKSPRQRTGIDAT
QVLHVKLRDRGNTVDFYHVTETLINITEOHIOQDPLREAYKRILEHYQRTGEIPR
DLRYLVRVEERELEELKGLERENKLSGDIRTKAMLTITMLLTPHPSLEDIRK
FYSLRAELKVKVGEESFLPVHGVDAEDVASSFAAIRKAPERERVVYRREISGFEPL
FHAAPHIKGVMSVLAATRSNIVSGEKHEEFIRHLERTQVRPLE"
complement(5360..5668)
/gene="APE0284"
complement(5360..5668)
/gene="APE0284"
/codon_start=1
/transl_table=11
/product="102aa long hypothetical protein"
/protein_id="BAA79238.1"
/db_xref="GI:5103922"
/transl_table="MSGTCGRRRVRSALPPARDPLVLCRGPFRGSCARRALPPIPFPLP
HEGLPLGVAGILLHDAASRPYRCRPECSNTDQTSVYIFSTSDHPCLLLLSPRFL"

Query Match          9.2%; Score 43.4; DB 1; Length 167000;
Best Local Similarity 48.3%; Pred. No. 0.13;
Matches 153; Conservative 0; Mismatches 161; Indels 3; Gaps 1;
OY 115 GGGAATTAATCGAGCTGGAAAGGAAGAAAGTCGTTATTCTCCCAAGGAATACGCC 174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38141 GCGGAGTTCTACGAATCGGTGAGGCCACGAGATTATAGTCGGGCCGGGAGCAGCATG 38200
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 175 TTAATCCTAACCCTCGACAGGATAAAGTTGCCCGACGATGTTATGGGGGATATGAAGATA 234

```


VERSION	AE001529.1	GI:4155590	gene	1563..2243
KEYWORDS	Helicobacter pylori J99.		CDS	/gene="jhp1007"
SOURCE	Helicobacter pylori J99			1563..2243
ORGANISM	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.			/gene="jhp1007"
REFERENCE	1 (bases 1 to 10922)			/note="similar to H. pylori 26695 gene HP0374"
AUTHORS	Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F., and Trust, T.J.			/codon_start=1
TITLE	Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori			/transl_table=1
JOURNAL	Nature 397 (6715), 176-180 (1999)			/product="putative"
MEDLINE	99120557			/protein_id="AAD06587.1"
REMARK	Erratum: [[published erratum appears in Nature 1999 Feb 25; 397(6721):7191]]			/db_xref="GI:4155595"
REFERENCE	2 (bases 1 to 10922)			FLYVYHAEIAKHALRLVOCPLIEIMANKKTHLILSVLEIKKXILKPLNLKDG
AUTHORS	King, B.L., Alm, R.A., and Trust, T.J.			KLISLFDADFQORNEKIDSKLEROKILIHSCQCGSGLMELFAFNTKALAYPK
TITLE	Direct Submission			ASVLDFNGEILPASADPEKGVILGPEGFSQERGYKEREIYRIPLDMLKSEACV
JOURNAL	Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA			FVASTA01"
COMMENT	Address all correspondence to: hp@arch.us.astro.com or Richard A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith, Bryndon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills, Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and Diane E. Taylor are with the University of Alberta Department of Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G 2H7 and the Canadian Bacterial Diseases Network. All other authors are with Astra Research Center Boston, 128 Sidney Street, Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capabilities are available at ARCB's world wide web site. (URL: http://www.astro-boston.com/hpylori).			complement(2327..4582)
FEATURES	Location/Qualifiers			complement(2327..4582)
source	1..10922			/note="similar to H. pylori 26695 gene HP0373"
gene	/organism="Helicobacter pylori J99"			/codon_start=1
CDS	/strain="J99"			/transl_table=1
	/db_xref="taxon:85963"			/product="putative outer membrane protein"
	complement(68..1075)			/protein_id="AAD06586.1"
	/gene="hemh"			/db_xref="GI:4155594"
	/note="jhp1005"			/translation="MLRLVSKTICLSLIFNLEAFQKKHOKDVFVEAGPETGLLEG
	complement(68..1075)			AQTEQAIAQNTQNTKIYENPLTHPQTEQPKQKSDTATPOSYGRYYILQNTIL
	/gene="hemh"			EKAPELTAAININGLTYFSONPVYMYAKDNEAFEGYGNNSVYVIONPLPYNLNN
	/note="similar to H. pylori 26695 gene HP0376"			IELSYDAQKAVNLGVIETPKDSQIILPASLFNPFNSDPSNDGLQQLQTTTFF
				SDANTOSLFELKSLQITTLQMTYENTDFSGNNDPNGLASPKPHYECFQYKKSQV
				ASVSFTQTAELTNLMIDIAVDFSKSWEAVLNAPQFNSPSECGIDYPKCNPF
				NGVLGDPKDKYALTPPEVINSRVANELTVNLLAAKGLGSLGNSMAYQVRLLQ
				QNGVLGDPKALDPKELKRNKVALALRDIIEYCHTLGTHNGNNTYQVRLLQ
				EGNGPEARCGGHEVKNRELEFSGHVEPHVDHGYTDVCSFRGNKQNPATPSNP
				NSITNCAQVPAVLGIVTAVQQLINQNALPINFANLSQTSHLNAGLNAQNFATSM
				VSAIAQNFSTSTTYSSSKNFRSLPVGNVKGIOHYFNDYIGLAYYGIQVNYAQ
				ANDEKIQOGLSYGGMDVLFDTITTKKQDHPKTKVFASFSGVFGGLRGLNYSYVF
				NOVKGSGNLDIVTGFNRYKHSYSGVSLVSIQSGIKIASNNNGIYADSVVLNKGSH
				FKVFNFYGVWF"
				complement(4751..5317)
	/gene="dcd"			/gene="dcd"
	/note="jhp1009"			/note="jhp1009"
	complement(4751..5317)			complement(4751..5317)
	/gene="dcd"			/gene="dcd"
	/note="similar to H. pylori 26695 gene HP0372"			/note="similar to H. pylori 26695 gene HP0372"
				/codon_start=1
	/transl_table=1			/transl_table=1
	/product="DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE"			/product="DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE"
	/protein_id="AAD06585.1"			/protein_id="AAD06585.1"
	/db_xref="GI:4155593"			/db_xref="GI:4155593"
	/translation="MGLKADSWIKKMSLEHGMNPFCEKQKGVNIVSYGLSSYGYDIR			/translation="MGLKADSWIKKMSLEHGMNPFCEKQKGVNIVSYGLSSYGYDIR
	VGSEFMFLDNKALIDKEDPNNAKIDASKEGFLLPANAFALANTIEYKPKPKDT			VGSEFMFLDNKALIDKEDPNNAKIDASKEGFLLPANAFALANTIEYKPKPKDT
	LVICLGSTVARGCIIYVNPFEPEPEGYITIEISNTNLPKAVYANEGIAQVVFLOG			LVICLGSTVARGCIIYVNPFEPEPEGYITIEISNTNLPKAVYANEGIAQVVFLOG
	DEVCEQSKYKONGRQIQGVGIIILPKILR"			DEVCEQSKYKONGRQIQGVGIIILPKILR"
	5447..5932			5447..5932
	/gene="accb"			/gene="accb"
	/note="jhp1010"			/note="jhp1010"
	5447..5932			5447..5932
	/gene="accb"			/gene="accb"
	/note="similar to H. pylori 26695 gene HP0371"			/note="similar to H. pylori 26695 gene HP0371"
				/codon_start=1
	/transl_table=1			/transl_table=1
	/product="BIOTIN CARBOXYL CARRIER PROTEIN"			/product="BIOTIN CARBOXYL CARRIER PROTEIN"
	/protein_id="AAD06584.1"			/protein_id="AAD06584.1"
	/db_xref="GI:4155592"			/db_xref="GI:4155592"
	/translation="MNLSEIEELIKFEKASDLGLKHLKHEHFLVDKESAYTKHKNA			/translation="MNLSEIEELIKFEKASDLGLKHLKHEHFLVDKESAYTKHKNA
	LNSHSSAPIQAPIMVEASIPSVQTPVPMVCTPIVDKKEDEVLSVPMGTYHAPSPGA			LNSHSSAPIQAPIMVEASIPSVQTPVPMVCTPIVDKKEDEVLSVPMGTYHAPSPGA
	EPYVKGVDTLKKGQIIGIVFAMKIMNEIEYPCVKVVSIEVGDAPQVPEYGTKLKVEK			EPYVKGVDTLKKGQIIGIVFAMKIMNEIEYPCVKVVSIEVGDAPQVPEYGTKLKVEK
	L"			L"
	5938..7305			5938..7305
	/gene="accb"			/gene="accb"
	/note="jhp1011"			/note="jhp1011"
	5938..7305			5938..7305

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 11:29:18 ; Search time 1175.76 Seconds
(without alignments)
4304.672 Million cell updates/sec

Title: US-08-957-709-70
Perfect score: 471
Sequence: 1 ATGCTACTTCAGACTGAA.....CAAGAGAGAAAGAACTCTAG 471

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	4.5	577	13 B59570	B59570 CIT-HSP-342
2	21	4.5	592	11 BF041541	BF041541 Bp250026A
3	21	4.5	697	10 A1982253	A1982253 pat.pk007
4	20	4.2	222	10 AV253548	AV253548 AV253548
5	20	4.2	413	11 BF551925	BF551925 UI-R-C2p-
6	20	4.2	517	10 BE503276	BE503276 7a15a08.x
7	20	4.2	530	13 A0983242	A0983242 RPCI-23-3
8	20	4.2	1048	12 AK005764	AK005764 Mus muscu
9	20	4.2	1172	13 CNS04VFP	AL309022 Tetraodon
10	20	4.2	1504	12 AK015811	AK015811 Mus muscu
11	19	4.0	240	10 BE217438	BE217438 fbn-i080
12	19	4.0	279	10 BB497535	BB497535 BB497535

C 13	19	4.0	310	13	AZ835592
C 14	19	4.0	349	10	AU221733
C 15	19	4.0	374	10	AV760824
C 16	19	4.0	394	10	AA494240
C 17	19	4.0	402	11	R07311
C 18	19	4.0	411	11	T79710
C 19	19	4.0	412	13	AQ895639
C 20	19	4.0	414	11	T77230
C 21	19	4.0	422	13	AQ280008
C 22	19	4.0	449	13	AQ083417
C 23	19	4.0	454	10	AA783647
C 24	19	4.0	460	10	BE233358
C 25	19	4.0	654	10	AU055942
C 26	19	4.0	673	13	AG017560
C 27	19	4.0	679	13	AZ851196
C 28	19	4.0	753	10	AV709981
C 29	19	4.0	762	13	AQ738645
C 30	19	4.0	770	13	BH068355
C 31	19	4.0	820	13	BH059058
C 32	19	4.0	835	11	BG340930
C 33	19	4.0	851	10	BE260608
C 34	19	4.0	1051	11	BI102504
C 35	18	3.8	200	11	Z40844
C 36	18	3.8	216	10	AV282403
C 37	18	3.8	246	10	AV172012
C 38	18	3.8	268	10	AV045201
C 39	18	3.8	268	10	AV047443
C 40	18	3.8	270	10	BB331361
C 41	18	3.8	277	11	BG378684
C 42	18	3.8	284	10	AL385496
C 43	18	3.8	286	10	AW713560
C 44	18	3.8	296	10	AV300832
C 45	18	3.8	304	10	AW526526

ALIGNMENTS

RESULT 1
B59570
LOCUS CIT-HSP-342G11.TPB CIT-HSP Homo sapiens genomic clone 342G11, DNA
DEFINITION B59570
ACCESSION B59570
VERSION B59570.1
KEYWORDS GI:2614288
SOURCE GSS.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 577)
JOURNAL Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden ,K., Bently,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.
COMMENT Use of a random BAC End Sequence Database for Sequence-Ready Map Building
Other_GSSs: CIT-HSP-342G11.TVB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdaams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..577
/organism="Homo sapiens"

```

/db_xref="GDB:5362141"
/db_xref="taxon:9606"
/clone="342G11"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT      222 a   76 c   84 g   195 t
ORIGIN

Query Match      4.5%; Score 21; DB 13; Length 577;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  11 CAGACTGGAATAATCAGAAAG 31
      |||||
Db   301 CAGACTGGAATAATCAGAAAG 321

RESULT 2
LOCUS      BF041541      592 bp      mRNA      EST      10-OCT-2000
DEFINITION clone BP250026A10A11 5', mRNA sequence.
ACCESSION  BF041541
VERSION     BF041541.1 GI:10758596
KEYWORDS   EST.
SOURCE     cow, taurus
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
            1 (bases 1 to 592)
REFERENCE  Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
            J.H.
            Bovine ESTs
            Unpublished
            Contact: Lewin, H. A.
            W. M. Keck Center for Comparative and Functional Genomics
            University of Illinois at Urbana-Champaign
            340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
            61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Funding for cattle EST sequencing was provided by the USDA National
            Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
            to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
            from Washington University Genome Center. Vector Trimmi g:
            Cross_match from Washington University Genome Center PHRAP suite.
            This sequence is vector free and at least 200 bp in length.
            PCR Primers
            FORWARD: TAATACGACTCACTATAGG
            BACKWARD: ATTAACCTCTACTAAG
            Insert Length: 592 Std Error: 0.00
            Plate: BP250026A10 row: A column: 11
            Seq primer: AGCGATAACAATTCACACGGA
            High quality sequence stop: 592.
            Location/Qualifiers
            1..592
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone="BP250026A10A11"
            /clone_lib="Soares normalized bovine placenta"
            /sex="female"
            /lab_host="DHI08"
            /note="Organ: Placenta; Vector: pT7T3Pac; Site_1: EcoRI;
            Site_2: NotI; The cDNA library was contributed by the
            Soares laboratory and it was constructed and normalized
            as described by Bonaldo, M.F., Lennon, G. and Soares,
            M.B. (1996), Genome Research 6(9): 791-806."

```

```

BASE COUNT      176 a   121 c   115 g   177 t   3 others
ORIGIN

Query Match      4.5%; Score 21; DB 11; Length 592;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  22 ATCAGAAAGAAATACATTATA 42
      |||||
Db   269 ATCAGAAAGAAATACATTATA 289

RESULT 3
LOCUS      AI982253      697 bp      mRNA      EST      07-MAY-2001
DEFINITION pat.pk0078.g2.f chicken activated T cell cDNA Gallus gallus CDNA
            O-PALMITOYLTRANSFERASE II, mRNA sequence.
ACCESSION  AI982253
VERSION     AI982253.1 GI:5885281
KEYWORDS   EST.
SOURCE     chicken.
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Arcosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
            1 (bases 1 to 697)
REFERENCE  Tirunaguru,V.G., Sofer,L., Cui,J. and Burnside,J.
            An expressed sequence tag database of T-cell-enriched activated
            chicken splenocytes: sequence analysis of 5251 clones
            Genomics 66 (2), 144-151 (2000)
            Contact: Joan Burnside
            Molecular Endocrinology
            University of Delaware
            40 Townsend Hall, Newark, DE 19717, USA
            Tel: 302 831-1345
            Fax: 302-831-3411
            Email: joan@udel.edu, www.chickest.udel.edu
            Seq primer: T7.
            Location/Qualifiers
            1..697
            /organism="Gallus gallus"
            /db_xref="taxon:9031"
            /clone="pat.pk0078.g2.f"
            /clone_lib="Chicken activated T cell cDNA"
            /sex="male"
            /cell_type="Con A-activated splenic T cell"
            /lab_host="E.coli TOP10 F'"
            /note="Vector: pCDNA3"

BASE COUNT      206 a   148 c   154 g   168 t   21 others
ORIGIN

Query Match      4.5%; Score 21; DB 10; Length 697;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  100 GAGGCTTTTGTAAAGGGAAA 120
      |||||
Db   470 GAGGCTTTTGTAAAGGGAAA 490

RESULT 4
LOCUS      AV253548      222 bp      mRNA      EST      04-NOV-1999
DEFINITION musculus cDNA clone 4921501122 3', mRNA sequence.
ACCESSION  AV253548
VERSION     AV253548.1 GI:6241007
KEYWORDS   EST.
SOURCE     house mouse.

```



```

Query Match      4.2%; Score 20; DB 11; Length 413;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 AAATCAGAAAGAAACTT 39
|||||
Db 71 AAATCAGAAAGAAACTT 90
|||||

RESULT 6
LOCUS BE503276/c 517 bp mRNA EST 04-AUG-2000
DEFINITION 7a15a08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3218774 3',
mRNA sequence.
ACCESSION BE503276
VERSION BE503276
KEYWORDS EST.
SOURCE EST.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 517)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400p from Gibco
High quality sequence stop: 463.

FEATURES
Location/Qualifiers
1..517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3218774"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following MAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 89 a 157 c 137 g 134 t
ORIGIN

Query Match      4.2%; Score 20; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TCAGAGTGGCAGAGGCT 105
|||||
Db 176 TCAGAGTGGCAGAGGCT 157
|||||

RESULT 7
LOCUS AQ983242 530 bp DNA GSS 30-JAN-2000
DEFINITION RPCI-23-307F24.TJ RPCI-23 Mus musculus genomic clone RPCI-23-307F24
, DNA sequence.
ACCESSION AQ983242
VERSION AQ983242.1 GI:6816447
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 530)
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-23-307F24.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Reseach Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 307 row: F column: 24
Seq primer: SP6
Class: BAC ends.

FEATURES
Location/Qualifiers
1..530
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-307F24"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 172 a 81 c 95 g 182 t
ORIGIN

Query Match      4.2%; Score 20; DB 13; Length 530;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 GATATGGAGAGAGATTGTC 369
|||||
Db 360 GATATGGAGAGAGATTGTC 379
|||||

RESULT 8
LOCUS AK005764 1048 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:1700008F21, full insert sequence.
ACCESSION AK005764
VERSION AK005764.1 GI:12838506
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:1700008F21.

```


1

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 240)
Schwerin, M., Dorroch, U. and Goldammer, T.
Mapping of differentially expressed hepatic and intestinal EST's in
cattle
JOURNAL Unpublished (2000)
AUTHORS Contact: Schwerin M
TITLE Molecular Biology
COMMENT PEN Dummerstorf
2nd Wilhelm-Stahl-Allee, Dummerstorf, 18196, Germany
Tel: 0049-038208-68700
Fax: 0049-038208-68702
Email: schwerin@fn-dummerstorf.de
High quality sequence stop: 240
POLYA=Yes.

FEATURES
source
1. .240
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Cattle puc18 Library"
/sex="female"
/tissue_type="Intestine"
/dev_stage="lactation"
/lab_host="X11-blue"
/note="Vector: pUC18 Sma I/BAP"
BASE COUNT 96 a 24 c 60 g 60 t
ORIGIN

Query Match 4.0%; Score 19; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 132 GGAAGAGGAGGAAAGTC 150
|||||
Db 156 GGAAGAGGAGGAAAGTC 174

RESULT 12
BB497535 . 279 bp mRNA EST 26-JUL-2000
LOCUS BB497535 RIKEN full-length enriched, 0 day neonate kidney Mus
DEFINITION musculus cDNA clone D630009E21 3', mRNA sequence.
BB497535
EST.
BB497535.1 GI:9471577
house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 279)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurlinara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata
Y., Shigenoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
N. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
Location/Qualifiers
1. .279
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10b"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCAACTCGAGTTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCTCGATTAAATTAATTAATCCCGCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 84 a 58 c 42 g 95 t
ORIGIN

Query Match 4.0%; Score 19; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 404 GAAACCTTACAGAGGAAA 422
|||||
Db 5 GAAACCTTACAGAGGAAA 23

RESULT 13
AZ835592/c 310 bp DNA GSS 20-FEB-2001
LOCUS 2M0129F24R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC2M0129F24 R, DNA sequence.
AZ835592
ACCESSION AZ835592
VERSION AZ835592.1 GI:13005500
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 310)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright,D.,Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0129 row: F column: 24
 Seq primer: CACACAGGAACACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 310.
 Location/Qualifiers
 1
 Location:Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2N0129F24"
 /clone_lib="Mouse 10kb plasmid UUGCLM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 58 a 80 c 68 g 104 t

Query Match 4.0%; Score 19; DB 13; Length 310;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TGGAAATCAGAAAGAAA 34
 Db 263 TGGAAATCAGAAAGAAA 245

RESULT 14
 AU221733/c 349 bp mRNA EST
 LOCUS AU221733 unpublished oligo-capped cDNA library, stage L1
 DEFINITION Caenorhabditis elegans cDNA clone yk897e04 3', mRNA sequence.
 ACCESSION AU221733
 VERSION AU221733.1 GI:14859890
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
 1; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 349)
 Kohara,Y., Shin-I.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
 and Sugano,S.

A complementary view of the C.elegans genome
 Unpublished (2001)
 Contact: Yuji Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.
 Location/Qualifiers
 1 349
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /db_xref="yk897e04"
 /clone_lib="unpublished oligo-capped cDNA library, stage L1"
 /sex="Hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 BASE COUNT 70 a 102 c 80 g 95 t 2 others

Query Match 4.0%; Score 19; DB 10; Length 349;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 AGCGTTTCAAGAGAAAG 462
 Db 185 AGCGTTTCAAGAGAAAG 167

RESULT 15
 AV760824/c 374 bp mRNA EST
 LOCUS AV760824 MDS Homo sapiens cDNA clone MDEDE05 5', mRNA sequence.
 DEFINITION AV760824
 ACCESSION AV760824.1 GI:10918672
 VERSION EST.
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 374)
 Gu,J., Zhao,M., Huang,Q., Xu,X., Li,Y., Peng,Y., Song,H., Xiao,H.,
 Gu,Y., Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng
 G., Xu,S., Gu,W., Yu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G.,
 Yang,Y., Gao,G., Zhang,Q., Chen,S., Han,Z. and Chen,Z.
 Homo sapiens cDNA MDS clones
 Unpublished (2000)
 Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn
 This clone is available at CHGC in Shanghai.
 Location/Qualifiers
 1 374
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MDEDE05"
 /clone_lib="MDS"
 /tissue_type="Bone marrow"
 /cell_type="CD34+ hematopoietic stem/progenitor cell"
 /lab_host="BM25, 8"
 /note="Vector: pTriplex2; Site_1: sf11A; Site_2: sf11B"

BASE COUNT 121 a 75 c 82 g 96 t

Query Match	4.0%	Score 19;	DB 10;	Length 374;
Best Local Similarity	100.0%;	Pred. No. 65;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	449	TTTCAAGAGAGAAAGAACT	467	
Db	88	TTTCAAGAGAGAAAGAACT	70	

Search completed: January 31, 2002, 11:49:22
Job time: 1204 sec

This Page Blank (uspto)

OM of: US-08-957-709-70 to: PIR_68:* out_format : pfs
Date: Jan 31, 2002 12:59 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE1=frame*_n2p.model -DEV=xih
-Q/cgn2.1/USPTO.spool/US08957709/runat_30012002.133258.9503/app_query.fasta_1.536
-DB=PIR_68 -OFMT=fastan -SUFFIX=rdr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DEPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-FDEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEADSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US08957709_@cgn1_1.67
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLIFY -WAIT -THREADS=1

Search information block:

Query: US-08-957-709-70
Query length: 471
Database: PIR_68:*
Database sequences: 219241
Database length: 7617452
Search time (sec): 39.880000

score_list:

Sequence	Strd	Orig	zscore	EScore	Len	Documentation
pir2:E71216	+	768.00	1427.63	1.0e-71	156	dCTP deaminase (EC 3.5.4.13) PH
pir2:G75030	+	703.00	1306.40	5.8e-65	154	dCTP deaminase (EC 3.5.4.13) PH
pir2:A70439	+	243.00	446.12	4.1e-17	180	probable dCTP deaminase (EC 3.5.4.13) PH
pir2:H83695	+	242.50	445.34	4.6e-17	177	deoxycytidine triphosphate deam
pir2:S26382	+	243.00	444.62	5.2e-17	173	probable dCTP deaminase (EC 3.5.4.13) PH
pir2:D72724	+	235.00	431.13	2.8e-16	181	probable dCTP deaminase (EC 3.5.4.13) PH
pir2:T36613	+	216.00	395.16	2.7e-14	191	probable dCTP deaminase (EC 3.5.4.13) PH
pir2:A69114	+	199.00	362.99	1.6e-12	200	dCTP deaminase (EC 3.5.4.13) PH
pir2:B70526	+	198.00	361.60	2.0e-12	190	dCTP deaminase (EC 3.5.4.13) PH
pir2:A81272	+	189.50	345.93	1.5e-11	186	probable dCTP deaminase (EC 3.5.4.13) PH
pir2:G84184	+	184.50	336.16	5.0e-11	195	deoxycytidine triphosphate deam
pir1:D64566	+	181.00	329.86	1.2e-10	190	dCTP deaminase (EC 3.5.4.13) PH
pir2:E71860	+	180.00	328.09	1.5e-10	188	dCTP deaminase (EC 3.5.4.13) PH
pir2:C84942	+	174.00	316.04	6.3e-10	206	dCTP deaminase (EC 3.5.4.13) PH
pir2:A83210	+	172.00	313.16	1.0e-09	188	probable deoxycytidine triphosph
pir2:C85833	+	165.00	299.85	5.4e-09	193	2'-deoxycytidine 5'-triphosphat
pir2:G82765	+	164.50	299.01	6.0e-09	191	deoxycytidine triphosphate deam
pir1:A42940	+	163.00	296.11	8.7e-09	193	dCTP deaminase (EC 3.5.4.13) PH
pir2:H72759	+	159.00	290.21	2.2e-08	163	probable dCTP deaminase (EC 3.5.4.13) PH
pir2:D81149	+	155.00	281.42	5.9e-08	188	deoxycytidine triphosphate deam
pir2:F64333	+	154.00	278.79	7.6e-08	204	dCTP deaminase (EC 3.5.4.13) PH
pir2:D69081	+	153.00	279.78	9.1e-08	150	deoxycytidine 5-triphosphate nuc
pir2:D81717	+	150.50	272.92	1.7e-07	190	deoxycytidine triphosphate deam
pir2:B71565	+	149.50	271.05	2.2e-07	190	probable dCTP deaminase (EC 3.5.4.13) PH
pir1:A64050	+	149.50	270.81	2.2e-07	195	dCTP deaminase (EC 3.5.4.13) PH
pir2:E71715	+	145.00	262.75	6.4e-07	188	probable dCTP deaminase (EC 3.5.4.13) PH
pir2:F66539	+	138.00	249.58	3.4e-06	190	dCTP deaminase [imported] - Chl
pir2:F72084	+	138.00	249.58	3.4e-06	190	dCTP deaminase (EC 3.5.4.13) PH
pir2:T44356	+	137.00	248.61	4.3e-06	172	probable dCTP deaminase (EC 3.5.4.13) PH
pir1:S75598	+	135.50	244.77	6.3e-06	193	dCTP deaminase (EC 3.5.4.13) PH
pir2:C69388	+	126.00	228.32	5.9e-05	168	probable dCTP deaminase (EC 3.5.4.13) PH
pir1:GNLJEV	+	125.50	209.60	9.6e-05	1145	! pol polyprotein - equine infed
pir1:GNLJEV	+	125.50	209.59	9.6e-05	1146	! pol polyprotein (clone 1369) -
pir1:GNLJ22	+	125.50	209.59	9.6e-05	1146	! pol polyprotein - equine infed
pir3:JC7565	+	119.00	214.21	0.0003	188	nucleoside-triphosphate (EC 3.5.4.13) PH
pir2:E64437	+	118.50	214.71	0.0004	161	probable dCTP deaminase (EC 3.5.4.13) PH
pir1:B45345	+	118.00	195.89	0.0006	1109	! pol polyprotein - caprine arth
pir2:A46256	+	112.50	204.74	0.0015	141	! dUTP pyrophosphatase (EC 3.6.1.1)
pir2:G02777	+	112.50	203.34	0.0015	164	! dUTP pyrophosphatase (EC 3.6.1.1)
pir1:GNLJVS	+	112.00	184.72	0.0024	1101	! pol polyprotein - Maedi/Visna

pir1:B45390 + 112.00 184.76 0.0024 1101 ! pol polyprotein - Maedi/Vis
pir1:B46335 + 111.50 183.95 0.0027 1086 ! pol polyprotein - Maedi/Vis
pir2:J01162 + 108.50 178.34 0.0056 1087 ! Pol protein - Maedi/Visna v
pir2:S26429 + 107.50 193.24 0.0050 178 ! probable dUTP pyrophosphatas
pir2:F84406 + 103.00 185.54 0.0146 165 ! deoxycytidine triphosphate d
seq_name: pir2:E71216
seq documentation block:
dCTP deaminase (EC 3.5.4.13) PH1997 [similarity] - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: E71216
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137
A:Accession: E71216
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-156 <KAW>
A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31124.1; PID:g3258441
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenB
C:Genetics:
A:Gene: PH1997
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

alignment_scores:

Quality: 768.00 Length: 156
Ratio: 5.020 Gaps: 0
Percent Similarity: 98.077 Percent Identity: 93.590

alignment_block:

US-08-957-709-70 x E71216

Align seg 1/1 to: E71216 from: 1 to: 156

1 ATGCTACTTCAGACTGGAAATCAGAAAGAAATACCTATAGAGCCATT 50
|||||
1 MetLeuLeuProAspTrpLysileArgLysGluLeuLeuGluProPh 17
51 TTCTGAAGATCGCTCCACACAGCAGGTATGACCTCAGCTGGGCAGAG 100
|||||
17 eSerGluGluSerLeuGluProAlaGlyTyrAspLeuArgValGlyArgG 34
101 AGGCTTTTGTAAAGGGGAAATTAATCGACGTGGAAAGGAAAGAAAGTC 150
|||||
34 LuAlaPheValSerGlyLysLeuLeuLeuValGluLysVal 50
151 GTTATTCCTCAAGGGAATACGCCTTAATCTCAACCCCTCGAGAGGATAA 200
|||||
51 VallileProProArgGluTyrAlaLeuLeuLeuThrLeuGluArgVal 67
201 GTTCCCGCCAGCATGTTATGGGGGATATGAACATAGGAGCAGTTAGCAA 250
|||||
67 sLeuProAspAspValMetGlyAspMetLysileArgSerSerLeuAla 84
251 GAGAGGGGGTATTGGTCTTCTTCTGGGTGGGTGACCCAGGATGGGATGA 300
|||||
84 rgGluGlyLeuLeuGlySerPheAlaTrpValAspProGlyTrpAspGly 100
301 AACTTAACACTTAATGCTCTACAAATGCCTCAATGAACCTGTGCAATTAAG 350
|||||
101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 117
351 ATATGGAGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGGTCGG 400
|||||
117 styGlyGluArgPheValGlnIleValPheAlaLeuGluAspProp 134

401 CAAGAAACCCCTTACAGAGAACTATCAGGGAGACACAAAGGTTAGCGTTT 450
|||||
134 roArgAsnProTyrSerGlyAsnTyrGlnGlySerThrArgLeuValPhe 150
|||||
451 TCAAGAGAGAAAGAACTC 468
|||||
151 SerLysArgLysLysLeu 156

seq_name: pir2:G75030

seq_documentation_block:
dCTP deaminase (EC 3.5.4.13) PAB1164 [similarity] - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: G75030
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: G75030
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50685.1; PID:g545919
C:Genetics:
A:Experimental source: strain Orsay
A:Gene: dcd; PAB1164
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

alignment_scores:
Quality: 703.00 Length: 153
Ratio: 4.718 Gaps: 0
Percent Similarity: 97.386 Percent Identity: 85.621

alignment_block:
US-08-957-709-70 x G75030 ..

Align seg 1/1 to: G75030 from: 1 to: 154

1 ATGCTACTTCAGACTGGAAATCAGAAATACATACTATAGACCAT 50
|||||
1 MetLeuLeuProAspTrpLysIleArgLysGluIleLeuLysProPh 17
|||||
51 TTCTGAAGAATCGCTCCACACAGCAGGTTATGACCTCAGAGTGGCGAG 100
|||||
17 eSerGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyLysG 34
|||||
101 AGGCTTTTGTAAAGGGAAATTAATCGACGTGGAAAGGAAAGAAAGTC 150
|||||
34 luAlaTyrIleGlnGlySerPheIleAspValGluLysGluGlyVal 50
|||||
151 GTTATTCTCTCAAGGGAATACGCCCTTAATCTTAACCCCTCGAGAGATAA 200
|||||
51 IleIleProLysGluTyrAlaLeuIleLeuThrLeuGluArgIleLy 67
|||||
201 GTTGGCCGACGATGTTATGGGGATATCAAGATAAGGAGCAGTTAGCAA 250
|||||
67 sLeuProAspAspIleMetGlyAspMetLysIleArgSerLeuAla 84
|||||
251 GAGAGGGGTTATGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
|||||
84 rGluGlyValLeuGlySerPheAlaTrpValAspProGlyTyrPaspGly 100
|||||
301 AACTTAACACTAATGCTCTACAAAGCCCTCAATGAACCTGTCGAATTAAG 350
|||||
101 AsnLeuThrLeuMetLeuTyrAsnAlaSerGluLysGluValIleLeuAr 117
|||||
351 ATATGGAGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGGTCGG 400
|||||
117 gTyrLysGluArgPheValGlnIleAlaPheLeuArgLeuGluAlaProA 134

401 CAAGAAACCCCTTACAGAGAACTATCAGGGAGACACAAAGGTTAGCGTTT 450
|||||
134 laLysAsnProTyrArgGlyAsnTyrGlnGlySerArgArgIleValLeu 150
|||||
451 TCAAGAGAGA 459
|||||
151 SerLysArg 153

seq_name: pir2:A70439

seq_documentation_block:
probable dCTP deaminase (EC 3.5.4.13) dcd [similarity] - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C:Accession: A70439
R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70439
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-180 <RAF>
A:Cross-references: GB:AE000747; NID:g2983944; PIDN:AAC07499.1; PID:g2983951; GB:AE00
C:Genetics:
A:Experimental source: strain VF5
A:Gene: dcd
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

alignment_scores:
Quality: 243.00 Length: 162
Ratio: 2.229 Gaps: 5
Percent Similarity: 67.284 Percent Identity: 37.654

alignment_block:
US-08-957-709-70 x A70439 ..

Align seg 1/1 to: A70439 from: 1 to: 180

1 ATGCTACTTCAGACTGGAAATCAGAAATCAGAAATCAGAAATCAGAAAT 35
|||||
1 MetIleLeuSerAspArgSerIleArgGluLeuIleGluLysGlyGluLe 17
|||||
36 ACTTATAGAGCATTTCTGAAGAAATCGCTCCACACAGCAGGTTATGACC 85
|||||
17 uLysValGluProTyrGluProSerHisValGlnCysSerSerLeuAspL 34
|||||
86 TCAGAGTGGCGAGAGAG...GCTTTTGTAAAGGGGAAA...TTAATCGAC 129
|||||
34 euArgLeuGlyAsnGlnIleAlaLeuTyrGluGlyGluGlyValIleasp 50
|||||
130 GTGGAAGAGGAAGGAAAGTCGTT..... 153
|||||
51 ValLysLysGlyThrLysGlyValArgIleLeuGluPheGluGluTyrPh 67
|||||
154 ...ATTCTCTCAAGGGAATACGCCCTTAATCTTAACCTCGAGAGATAA 199
|||||
67 eAspIleMetProLysGlnPheLeuLeuAlaThrLeuGluLysIle 84
|||||
200 ACTTGGCCGACGATGTTATGGGGATATGAAGATAAGGAGCAGTTAGCA 249
|||||
84 erLeuProTyrValThrAlaPheValGluGlyArgSerSerLeuGly 100
|||||
250 AGAAGAGGGTT...ATTGGTTCTTTTCTTGGGTTGACCCAGCATGGGA 296
|||||
101 ArgLeuGlyLeuPheIleGluAsnAlaGlyTrpValAspAlaGlyPheGl 117
|||||
297 TGGAACTTAACACTAATGCTCTCAATGCTCAATGAACTGCTCGAAT 346
|||||
117 uGlyGlnIleThrLeuGluLeuPheAsnAlaAsnAspArgProIleArgL 134

alignment_scores:
Quality: 180.00 Length: 168
Ratio: 1.782 Gaps: 4
Percent Similarity: 60.119 Percent Identity: 27.976

alignment_block:

US-08-957-709-70 x E71860 ..

Align seg 1/1 to: E71860 from: 1 to: 188

```
37 CTTATAGACCATTTCTGGAAGA..... 60
   ::::||||| |||:
18 MetileAsnProPheCysGluLysValGlyLysAsnValIleSerTy 34
   ::::||||| |||:
61 .TCGCTCCCAACAGCAGGTATGACCTCAGAGTGGCGAGAG..... 102
   ::::||||| |||:
34 rGlyLeuSerTyGlyTyAspIleArgValGlySerGluPheMetL 51
   ::::||||| |||:
103 .....GCTTTGTTAAGGGA..... 120
   ::::||||| |||:
51 euPheAspAsnLysAsnAlaLeuIleAspProLysAsnPheAspProAsn 67
   ::::||||| |||:
121 .....TTAATCGAGCTGGAAGAAGAAAGTCGTTATTCCTCC 161
   ::::||||| |||:
68 AsnAlaThrLysIleAspIleSerLysGluGlyPhePheIleLeuProAl 84
   ::::||||| |||:
162 AAGGAAATACGCTTATCTTACCTCGAGAGTAAAGTTGCCGAGC 211
   ::::||||| |||:
84 asnAlaPheAlaLeuAlaHisThrIleGluTyPheLysMetProLysA 101
   ::::||||| |||:
212 ATGTATGGGGATATGAAGATAAGGACAGCTTAGCAAGAGAGGGTT 261
   ::::||||| |||:
101 spThrLeuAlaIleCysLeuGlyLysSerThrTyAlaArgCysGlyIle 117
   ::::||||| |||:
262 ATGTTGCTTTTGGTGGTGGTACCCAGGATGGATGGAACCTTAACACT 311
   ::::||||| |||:
118 IleValAsnValThrProPheGluProGluPheGluGlyTyIleThrI 134
   ::::||||| |||:
312 AATGCTCTACATGCCTCAATATGAACCTGTGCAATTAAGATATGGAGAGA 361
   ::::||||| |||:
134 eGluIleSerAsnThrAsnLeuProAlaLysValTyAlaAsnGluG 151
   ::::||||| |||:
362 GATTTGTGCAGATCGCAPTTATAAGGCTAGAGGTCGCGCAAGAACCT 411
   ::::||||| |||:
151 LyIleAlaGlnValValPheLeuGlnGlyAspGluValCysGluGlnSer 167
   ::::||||| |||:
412 TACAGA.....GGAACACTATCAGGGGAGCACAGGTTAGGCTTTC 452
   ::::||||| |||:
168 TyrLysAspArgGlyGlyLysTyGlnGlyGlnValGlyIleThrLeuPr 184
   ::::||||| |||:
453 AAAG 456
   |||
184 oLys 185
```

seq_name: pir2:C84942

seq_documentation_block:

dCTP deaminase (PC 3.5.4.13) [imported] - Buchnera sp. (strain APS)

N:Alternate names: deoxycytidine triphosphate deaminase

C:Species: Buchnera sp.

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001

C:Accession: C84942

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

Nature 407, 81-86, 2000

A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.

A:Reference number: A84930; MUID:20445173

A:Accession: C84942

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-206 <STO>

A:Cross-references: GB:AP000398; GSPDB:GN00144

A:Experimental source: strain APS

C:Genetics:

A:Gene: dcd; BU108
C:Keywords: hydrolase

alignment_scores:

Quality: 174.00 Length: 164
Ratio: 1.812 Gaps: 5
Percent Similarity: 58.537 Percent Identity: 29.268

alignment_block:

US-08-957-709-70 x C84942 ..

Align seg 1/1 to: C84942 from: 1 to: 206

```
7 CTTCCACAGTGAATAAAGAAATCTTATAGAGCCATTTCTGA 56
   ::::||||| |||:
21 IleGluGluTrpLeuGluArgLysGluLeuIleGluProTyProAs 37
   ::::||||| |||:
57 AGAATCCGTC...CAACCAGCAGGTATGACCTCAGAGTGGCGAGAG. 102
   ::::||||| |||:
37 nLysThrLeuIleAsnGlyIleThrValAspIleHisLeuGlyAsnLysP 54
   ::::||||| |||:
103 .....GCTTTGTTAAG.....GGGAAATTAATCGAGCTG. 132
   ::::||||| |||:
54 heArgPhePheTyGluHisThrGlySerCysIleAspLeuSerAsnSer 70
   ::::||||| |||:
132 ..... 132
71 LysIleIleGlyLeuSerLeuThrGluIleMetSerAsnGluIleI 87
   ::::||||| |||:
133 ....GAAAGGAAGAAAGTCGTTATTCCTCCAAAGGAATACGCTTAA 178
   ::::||||| |||:
87 ePheSerLysGluGlnProCysPheLeuGlnProGlySerLeuValLeu 104
   ::::||||| |||:
179 TCCTAACCCCTCGAGAGGATAAAGTTGCCGAGCATGTTATGGGGATATG 228
   ::::||||| |||:
104 yssrThrPheGluSerIleLysMetProAsnAsnLeuValGlyTrpLeu 120
   ::::||||| |||:
229 AAGATAAGGAGCAGCTTAGCAAGAGAGGGGTTATT.....GGTTCTTT 272
   ::::||||| |||:
121 AspGlyArgSerSerLeuAlaArgLeuGlyLeuMetIleHisAlaThrAl 137
   ::::||||| |||:
273 TGCCTGGGTTGACCCAGGATGGATGGAACCTTAACACTAATGCTCTACA 322
   ::::||||| |||:
137 aHisArgIleAspProGlyTrpAsnGlyAsnIleValLeuGluMetPheA 154
   ::::||||| |||:
323 ATGCCTCAAAATGAACCTGTGCAATTAAGATATGGAGAGATTTGTGCAG 372
   ::::||||| |||:
154 snAlaGlyLysLeuThrLeuValLeuArgProLysMetArgIleAlaAla 170
   ::::||||| |||:
373 ATCGCATTTATAAGGCTAGAGGTCGCGCAAGAAACCTTAC 414
   ::::||||| |||:
171 LeuSerPheGluValLeuSerGlnProValLeuArgProTy 184
```

seq_name: pir2:A83210

seq_documentation_block:

probable deoxycytidine triphosphate deaminase PA3480 [imported] - Pseudomonas aerugin

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: A83210

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: A83210

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-188 <STO>

A:Cross-references: GB:AE004769; GB:AE004091; MID:g9949624; PIDN:AAG06868.1; GSPDB:CN

A:Experimental source: strain PA01

C:Genetics:
A:Gene: PA3480
C:Superfamily: dCTP deaminase

```
alignment_scores:      Quality: 172.00      Length: 171
                       Ratio: 1.737        Gaps: 4
Percent Similarity: 57.895      Percent identity: 26.316

alignment_block:
US-08-957-709-70 x A83210  ..

Align seg 1/1 to: A83210 from: 1 to: 188

37  CTTATAGAGCCATTTCTGAGAATCGTCCAAACCAGCA..... 75
   ::::::::::::::::::::|
18  MetIleGluProPheValGluArgGlnValArgGlyAlaAspSerAr 34
   ::::::::::::::::::::|
76  .....GGTTATGACCTCAGAGTGGCA 97
   ::::::::::::::::::::|
34  gValIleSerTyrGlyValSerTyrGlyTyrAspValArgCysAlaA 51
   ::::::::::::::::::::|
98  GAGAG..... 102
   |||
51  IaGluPheLysValPheThrAsnIleHisSerAlaValValAspProLys 67
103 GCTTTTGTAAAGGGAATTAATCGACGTGGAAGGAAGAAAGTCTG 152
   ||| ::::::::::::::::::::|
68  AsnPheAspGluLysSerPheValAspIleAsnSerAsp...ValCysIl 83
153 TATTCCTCAAGGAATACGGCTTAATCCTAACCTCGAGAGATAAAGT 202
   ::::::::::::::::::::|
83  eIleProAsnSerPheAlaLeuAlaArgThrValGluTyrPheArgI 100
   ::::::::::::::::::::|
203 TGCCCGACAGTGTATGGGGGATATGAAGATAAGGAGCAGTTAGCAAGA 252
   ::| | | | | ::| | | | |
100  IeProArgAspValLeuThrIleCysLeuGlyLysSerThrTyrAlaArg 116
253 GAAGGGTTATGGTCTTTCTGGTGGTTCGACCCAGGATGGATGGAAA 302
   ||| ::| | | | | ::| | | | |
117  CysGlyIleIleValAsnValThrProLeuGluProGluTrpGluGlyHi 133
303 CTTAACACTAATGCTCTCAATGCTCAATGCACTGCTGAATTAAGAT 352
   ::| | | | | ::| | | | | | | | | | |
133  sValThrLeuGluPheSerAsnThrThrAsnLeuProAlaLysIleTyrA 150
353 ATGGAGAGAGATTTGTGAGATCGCATTTATAAGGCTAGAGGGTCCGGCA 402
   ::| | | | | ::| | | | | ::| | | | |
150  IaAsnGluGlyValAlaGlnMetLeuPheLeuGlnSerAspGluAlaCys 166
403 AGAAACCCCTTACAGA.....GGAACCTATCAGGGGAGCACAGGTT 443
   ::| | | | | | | | | | | | | | | |
167  GluValSerTyrLysAspArgGlyGlyLysTyrGlnGlyGlnArgGlyVa 183
444 AGCGTTTCAAAG 456
   ::| | |
183  IThrLeuProLys 187
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 12:14:38 ; Search time 38.96 seconds
(without alignments)
2737.966 Million cell updates/sec

Title: US-08-957-709-70
Perfect score: 471
Sequence: 1 ATGTACTTCCAGACTGAA.....CAAGAGAAAGAACTCTAG 471

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	316	67.1	740	US-08-822-774-42	Sequence 42, Appl
2	61	13.0	129	US-08-822-774-40	Sequence 40, Appl
3	45	9.6	84	US-08-822-774-36	Sequence 36, Appl
4	18	3.8	797	US-08-173-510B-101	Sequence 101, Appl
5	18	3.8	797	US-08-458-218-99	Sequence 99, Appl
6	18	3.8	797	US-08-450-497-101	Sequence 101, Appl
7	17	3.6	1197	US-08-861-269-4	Sequence 4, Appl
8	17	3.6	1197	US-09-134-596-4	Sequence 4, Appl
9	17	3.6	1197	US-09-293-273-4	Sequence 9, Appl
10	17	3.6	7399	US-08-418-848A-9	Sequence 9, Appl
11	17	3.6	9709	US-08-188-583-5	Sequence 5, Appl
12	17	3.6	9709	US-08-388-353-1	Sequence 1, Appl
13	17	3.6	9709	US-08-488-551B-1	Sequence 1, Appl
14	17	3.6	12494	US-08-935-312-13	Sequence 13, Appl
15	17	3.6	12494	US-08-848-760B-33	Sequence 33, Appl
16	17	3.6	15581	US-08-646-538-35	Sequence 35, Appl
17	17	3.6	15581	US-09-503-222-35	Sequence 35, Appl
18	16	3.4	60	US-09-046-247-80	Sequence 80, Appl
19	16	3.4	1181	US-09-329-796-2	Sequence 2, Appl
20	16	3.4	1533	US-07-721-761A-32	Sequence 32, Appl
21	16	3.4	1533	US-07-978-687-32	Sequence 32, Appl
22	16	3.4	1533	US-08-471-791-12	Sequence 12, Appl
23	16	3.4	1533	PCT-US91-01746-12	Sequence 12, Appl
24	16	3.4	1533	PCT-US91-05801-32	Sequence 32, Appl
25	16	3.4	1542	US-08-865-311-1	Sequence 1, Appl
26	16	3.4	1544	US-09-313-300-4	Sequence 4, Appl
27	16	3.4	2298	US-08-951-648-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-08-822-774-42
; Sequence 42, Application US/08822774

; Patent No. 6183997

; GENERAL INFORMATION:

; APPLICANT: HOGREFE, Holly

; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)

; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins, and Methods for Purifying and Identifying Same

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &

; ADDRESSEE: Lenahan, P.L.L.C.

; STREET: 1200 G Street, N.W. Suite 700

; CITY: Washington

; STATE: D.C.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/822,774

; FILING DATE: 21-MAR-1997

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: KULIK, David J.

; REGISTRATION NUMBER: 36,576

; REFERENCE/DOCKET NUMBER: 1486/43163

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 628-8800

; TELEFAX: (202) 628-8844

; INFORMATION FOR SEQ ID NO: 42:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 740 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-822-774-42

Query Match 67.1%; Score 316; DB 4; Length 740;
Best Local Similarity 100.0%; Pred.No. 2e-155;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 TCCTCCAAAGGAATACGCTTAATCCTTAACCCCTCGAGAGGATAAAGTTGCCCGACGATGT 215

```
Db 1 TCTCCAGGAATACGCCCTTAATCTTAACCCCTCGAGAGGATAAGTTGCCGACGATGT 60
QY 216 TATGGGGATATGAAGATAAGAGCAGTATTAGCAAGAGAGAGGGGTTATTGGTTCTTTTGC 275
Db 61 TATGGGGATATGAAGATAAGAGCAGTATTAGCAAGAGAGAGGGGTTATTGGTTCTTTTGC 120
QY 276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCCTCAATGA 335
Db 121 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTTAATGCTCTACAATGCCCTCAATGA 180
QY 336 ACCTGTCGAATTAAAGATATGAGAGAGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 395
Db 181 ACCTGTCGAATTAAAGATATGAGAGAGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGG 240
QY 396 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGCAGCAGGTTAGCGTTTTCAAA 455
Db 241 TCCGGCAAGAAACCCCTTACAGAGAACTATCAGGGAGCAGCAGGTTAGCGTTTTCAAA 300
QY 456 GAGAAAGAACTCTAG 471
Db 301 GAGAAAGAACTCTAG 316

RESULT 2
US-08-822-774-40/c
; Sequence 40, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-822-774-40
```

```
Query Match 13.0%; Score 61; DB 4; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGCTACTCCAGACTGGAAATCAGAAAAAATACTTATAGAGCCATTTCTGAAGAA 60
Db 98 ATGCTACTCCAGACTGGAAATCAGAAAAAATACTTATAGAGCCATTTCTGAAGAA 39
QY 61 T 61
Db 38 T 38

RESULT 3
US-08-822-774-36
; Sequence 36, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-822-774-36

Query Match 9.6%; Score 45; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 2.7e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTCTTAAGGGGAAATTAATCGACGTGAGGAGGAGGAGGAGGAGGAGGAGG 151
Db 21 TTCTTAAGGGGAAATTAATCGACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 65

RESULT 4
US-08-173-510B-101
; Sequence 101, Application US/08173510B
; Patent No. 5747296
; GENERAL INFORMATION:
; APPLICANT: MATTHEW MOYLE, ET AL.
; TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
```

```

RESULT      6
US-08-450-497-101
; sequence 101, Application US/08450497
; Patent No. 5919900
; GENERAL INFORMATION:
; APPLICANT: MATTHEW WOYLE, ET AL.
; TITLE OF INVENTION: NOVEL NEUTROPHIL INHIBITORS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,497
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMMR: 08/173,510

```

;; FILING DATE: 23-DEC-1993
;; APPLICATION NUMBER: 08/151,064
;; FILING DATE: 10-NOV-1993
;; APPLICATION NUMBER: 08/060,433
;; FILING DATE: 11-MAY-1993
;; APPLICATION NUMBER: 07/996,972
;; FILING DATE: 24-DEC-1992
;; APPLICATION NUMBER: 07/881,721
;; FILING DATE: 11-MAY-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BIGGS, SUZANNE L.
;; REGISTRATION NUMBER: 30,158
;; REFERENCE/DOCKET NUMBER: 205/073
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 955-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 101:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 797
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: SINGLE
;; TOPOLOGY: LINEAR
;; MOLECULE TYPE: NUCLEIC
;; FEATURE:
;; NAME/KEY: Coding Sequence
;; LOCATION: 3...725
;; US-08-450-497-101

Query Match 3.8%; Score 18; DB 2; Length 797;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 GTGAAAGGAGGAGAAA 147
|||||
DB 558 GTGAAAGGAGGAGAAA 575

RESULT 7
US-08-861-269-4
; Sequence 4, Application US/08861269
; Patent No. 5817494
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: UBIQUITIN CONJUGATION PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: Filed Herewith
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0302 US
; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 415-855-0555
;; TELEFAX: 415-845-4166
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1197 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: BRAINOT09
;; CLONE: 2151473
;; US-08-861-269-4

Query Match 3.6%; Score 17; DB 1; Length 1197;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GAAATCAGAAAGAGAAA 34
|||||
DB 834 GAAATCAGAAAGAGAAA 850

RESULT 8
US-09-134-596-4
; Sequence 4, Application US/09134596
; Patent No. 5923318
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: UBIQUITIN CONJUGATION PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,596
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/861,269
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0302 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT09
; CLONE: 2151473
; US-09-134-596-4

Query Match 3.6%; Score 17; DB 2; Length 1197;
Best Local Similarity 100.0%; Pred. No. 11;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GAAATCAGAAAGAAA 34
 |||||
 Db 834 GAAATCAGAAAGAAA 850

RESULT 9

US-09-293-273-4
 ; Sequence 4, Application US/09293273
 ; Patent No. 6057112
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Corley, Neil C.
 ; APPLICANT: Lal, Preeti
 ; APPLICANT: Shah, Purvi
 ; TITLE OF INVENTION: UBIQUITIN CONJUGATION PROTEINS
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/293,273
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/134,596
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0302 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1197 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: BRAINOT09
 ; CLONE: 2151473

US-09-293-273-4

Query Match 3.6%; Score 17; DB 3; Length 1197;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 GAAATCAGAAAGAAA 34
 |||||
 Db 834 GAAATCAGAAAGAAA 850

RESULT 10

US-08-418-848A-9
 ; Sequence 9, Application US/08418848A
 ; Patent No. 5847096
 ; GENERAL INFORMATION:
 ; APPLICANT: SCHUBERT, MANFRED, HARMISON II,
 ; APPLICANT: GEORGE G., CHANG-JIE, CHEN, BANUJEREA, AKHIL
 ; TITLE OF INVENTION: DEFECTIVE, INTERFERING
 ; HIV PARTICLES

NUMBER OF SEQUENCES: 77
 CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; COUNTRY: U.S.A.
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORD PERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/418,848A
 ; FILING DATE: 07-APR-1995
 ; CLASSIFICATION: 526
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/936,849
 ; FILING DATE: 28-AUG-1992
 ; CLASSIFICATION: 526
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: RICHARD W. BORK
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4091US2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-758-4800
 ; TELEFAX: 212-751-6849
 ; TELEX: 421792
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7399 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-418-848A-9

Query Match 3.6%; Score 17; DB 2; Length 7399;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAGGAGGAGAAA 147
 |||||
 Db 2632 TGGAAAGGAGGAGAAA 2648

RESULT 11

US-08-188-583-5
 ; Sequence 5, Application US/08188583
 ; Patent No. 5851813
 ; GENERAL INFORMATION:
 ; APPLICANT: Desrosiers, Ronald C.
 ; TITLE OF INVENTION: PRIMATE LENTIVIRUS VACCINES
 ; NUMBER OF SEQUENCES: 57
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: U.S.A.
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; COMPUTER: IBM PS/2 Model 502 or 55SX
 ; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 ; SOFTWARE: WordPerfect (Version 5.0)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/188,583
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/727,494
 ; FILING DATE: July 9, 1991

```

/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/551,945
/ FILING DATE: July 12, 1990
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Freeman, John W.
/ REGISTRATION NUMBER: Reg. No. 5851813 29.066
/ REFERENCE/DOCKET NUMBER: 00246/079002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 542-5070
/ TELEFAX: (617) 542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 9709
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-08-188-583-5

Query Match 3.6%; Score 17; DB 2; Length 9709;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAAGGAAGGAAAA 147
DB 2671 TGGAAAAGGAAGGAAAA 2687

RESULT 12
US-08-388-353-1
: Sequence 1, Application US/08388353
: Patent No. 6010895
: GENERAL INFORMATION:
: APPLICANT: Deacon, Nicholas J.
: APPLICANT: Learmont, Jennifer C.
: APPLICANT: McPhee, Dale A.
: APPLICANT: Crowe, Suzanne
: APPLICANT: Cooper, David
: TITLE OF INVENTION: NON-PATHOGENIC STRAINS OF HIV-1
: NUMBER OF SEQUENCES: 800
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Scully, Scott, Murphy & Presser
: STREET: 400 Garden City Plaza
: CITY: Garden City
: STATE: New York
: COUNTRY: United States
: ZIP: 11530
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/388,353
: FILING DATE: 14-FEB-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Digiglio, Frank S.
: REGISTRATION NUMBER: 31,346
: REFERENCE/DOCKET NUMBER: 9606
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (516) 742-4366
: TELEX: 230 901 SANS UR
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9709 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
/ US-08-388-353-1

Query Match 3.6%; Score 17; DB 3; Length 9709;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAAGGAAGGAAAA 147
DB 2671 TGGAAAAGGAAGGAAAA 2687

RESULT 13
US-08-488-551B-1
: Sequence 1, Application US/08488551B
: Patent No. 6015661
: GENERAL INFORMATION:
: APPLICANT: Nicholas J. Deacon
: APPLICANT: Dale A. McPhee
: APPLICANT: David Cooper
: TITLE OF INVENTION: NON-PATHOGENIC STRAINS OF HIV-1
: NUMBER OF SEQUENCES: 841
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
: STREET: 400 GARDEN CITY PLAZA
: CITY: GARDEN CITY
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 11530-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/488,551B
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PM3864 (AU)
: FILING DATE: 14-FEB-1994
: APPLICATION NUMBER: PM4002 (AU)
: FILING DATE: 21-FEB-1994
: APPLICATION NUMBER: PM0284 (AU)
: FILING DATE: 23-DEC-1994
: APPLICATION NUMBER: US 08/388,353
: FILING DATE: 14-FEB-1995
: APPLICATION NUMBER: PN3021/95
: FILING DATE: 17-MAY-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: FRANK S. DIGIGLIO
: REFERENCE/DOCKET NUMBER: 9606Z
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (516) 742-4366
: TELEFAX: (516) 742-4366
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9709 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
/ US-08-488-551B-1

Query Match 3.6%; Score 17; DB 3; Length 9709;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAAGGAAGGAAAA 147
DB 2671 TGGAAAAGGAAGGAAAA 2687

RESULT 14
US-08-488-551B-1
```

US-08-935-312-13
; Sequence 13, Application US/08935312
; Patent No. 6207455
; GENERAL INFORMATION:
; APPLICANT: CHANG, Lung-Ji
; TITLE OF INVENTION: LENTIVIRAL VECTORS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 624 Ninth Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935,312
; FILING DATE: 22-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: CHANG-112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12494 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
US-08-935-312-13

Query Match 3.6%; Score 17; DB 4; Length 12494;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAAGGAGGAAAA 147
|||||
DB 2342 TGGAAAAGGAGGAAAA 2358

RESULT 15
US-08-848-760B-33
; Sequence 33, Application US/08848760B
; Patent No. 6248721
; GENERAL INFORMATION:
; APPLICANT: Chang, Lung-Ji
; TITLE OF INVENTION: Animal Model For Evaluation Of Vaccines
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: United States of America
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/848,760B
; FILING DATE: 25-Jan-2001

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/838,702
FILING DATE: 09-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: PACE, DORAN R.
REGISTRATION NUMBER: 38,261
REFERENCE/DOCKET NUMBER: CNG-100C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 12494 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-08-848-760B-33

Query Match 3.6%; Score 17; DB 4; Length 12494;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAAGGAGGAAAA 147
|||||
DB 2342 TGGAAAAGGAGGAAAA 2358

Search completed: January 31, 2002, 12:15:46
Job time: 68 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 12:15:48 ; Search time 101.6 seconds

(without alignments)
3974.411 Million cell updates/sec

Title: US-08-957-709-70

Perfect score: 471

Sequence: 1 ATGCTACTTCAGACTGAA.....CAAGAGAAAGAACTCTAG 471

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	471	100.0	471	19	AAV63860 Polymerase enhanc
2	21	4.5	33	19	AAV63865 Polymerase enhanc
3	18	3.8	405	22	AAH8082 Peppermint plant o
4	18	3.8	571	22	AAH09366 Human cDNA clone (
5	18	3.8	638	22	AAH88046 Peppermint plant o
6	18	3.8	797	20	AAH85547 Canine hookworm ne
7	18	3.8	878	21	AAC59749 Human secreted pro
8	18	3.8	1057	22	AAF33251 Human secreted pro
9	18	3.8	2691	17	AAT12325 DNA encoding Sulfo
10	18	3.8	3137	22	AAH17117 Human cDNA sequenc
11	18	3.8	10723	12	AAQ12787 Dengue 2 virus gen

c	12	17	3.6	285	22	AAH50830	Human tumour assoc
c	13	17	3.6	309	22	AAH50736	Human tumour assoc
	14	17	3.6	382	8	AAH71175	Sequence encoding
	15	17	3.6	383	8	AAH70530	Sequence encoding
	16	17	3.6	788	21	AAH14761	Aspergillus oryzae
	17	17	3.6	819	22	AAH51904	HIV protease and r
	18	17	3.6	819	22	AAH51905	HIV protease and r
c	19	17	3.6	837	20	AAZ34491	Human LYST Interac
	20	17	3.6	840	21	AAH79698	Eucalyptus grandis
	21	17	3.6	955	22	AAH63826	Human polynucleoti
	22	17	3.6	1045	22	AAH51832	HIV protease and r
	23	17	3.6	1053	22	AAH51871	HIV protease and r
	24	17	3.6	1115	22	AAH51864	HIV protease and r
	25	17	3.6	1115	22	AAH51920	HIV protease and r
	26	17	3.6	1116	22	AAH51841	HIV protease and r
	27	17	3.6	1116	22	AAH51842	HIV protease and r
	28	17	3.6	1116	22	AAH51844	HIV protease and r
	29	17	3.6	1116	22	AAH51851	HIV protease and r
	30	17	3.6	1116	22	AAH51861	HIV protease and r
	31	17	3.6	1116	22	AAH51873	HIV protease and r
	32	17	3.6	1116	22	AAH51874	HIV protease and r
	33	17	3.6	1116	22	AAH51875	HIV protease and r
	34	17	3.6	1116	22	AAH51876	HIV protease and r
	35	17	3.6	1116	22	AAH51879	HIV protease and r
	36	17	3.6	1116	22	AAH51889	HIV protease and r
	37	17	3.6	1116	22	AAH51892	HIV protease and r
	38	17	3.6	1116	22	AAH51894	HIV protease and r
	39	17	3.6	1116	22	AAH51895	HIV protease and r
	40	17	3.6	1116	22	AAH51910	HIV protease and r
	41	17	3.6	1116	22	AAH51912	HIV protease and r
	42	17	3.6	1116	22	AAH51917	HIV protease and r
	43	17	3.6	1116	22	AAH51924	HIV protease and r
	44	17	3.6	1116	22	AAH51933	HIV protease and r
	45	17	3.6	1116	22	AAH51935	HIV protease and r

ALIGNMENTS

RESULT 1
AAV63860
ID AAV63860 standard; DNA; 471 BP.
XX
AC AAV63860;
XX
DT 01-MAR-1999 (first entry)
XX
DE Polymerase enhancing factor P45 (dUTPase) component DNA.
XX
KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;
KW amplification; sequencing; replication; ss.
XX
OS Pyrococcus furiosus strain DSM 3638.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX (STRA-) STRATAGENE.
PA Hansen CJ, Hogrefe H;
XX WPI: 1998-542284/46.
DR P-PSDB; AAW72847.
XX
PT Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication

XX PS Claim 17: Page 43; 161pp; English.

XX CC This DNA sequence encodes the P45 component (see AAW72847) of the

CC polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.

CC PCR primers (see AAV63861-64) based on an N-terminal peptide (see

CC AAW72846) of P45 were used to amplify *P. furiosus* genomic DNA, and

CC sequencing was carried out on purified PCR products and plasmid

CC mini-preps. P45 and P40 (see AAW72844) are the predominant components

CC of PEF, which acts to enhance the activity of *P. furiosus* DNA

CC polymerase, thereby providing replication products of greater

CC length and purity. P45 was identified as a dUTPase, and possesses

CC polymerase enhancing activity. The invention provides novel

CC extracts, proteins and complexes that improve the polymerisation

CC activity of nucleic acid polymerases, as well as DNA constructs and

CC antibodies. Also included are methods for identifying compositions

CC with polymerase enhancing activity, for purifying and using these

CC compositions, and specific extracts, proteins and complexes that

CC function to enhance polymerase activity. Nucleic acid polymerase

CC reactions can be enhanced (claimed) by mixing a nucleic acid

CC template, at least 1 polymerase and a composition having polymerase

CC enhancing activity. Kits are provided for replicating nucleic

CC acids. The kits can be used in site-directed mutagenesis, nucleic

CC acid sequencing or amplification (preferably PCR or RT-PCR).

CC Isolated DNA can be used in the recombinant production of P45.

XX SQ Sequence 471 BP; 156 A; 80 C; 124 G; 111 T; 0 other;

Query Match 100.0%; Score 471; DB 19; Length 471;

Best Local Similarity 100.0%; Pred. No. 3e-230; Indels 0; Gaps 0;

Matches 471; Conservative 0; Mismatches 0;

QY 1 ATGCTACTTCCAGCTGGAATACAGAAATACTTATAGAGCCATTTCTGAAGAA 60

DB 1 ATGCTACTTCCAGCTGGAATACAGAAATACTTATAGAGCCATTTCTGAAGAA 60

QY 61 TCCTCTCAACACAGCTGTTATGCTTACAGTGGCGAGAGCGCTTTTGAAGGGGAAA 120

DB 61 TCCTCTCAACACAGCTGTTATGCTTACAGTGGCGAGAGCGCTTTTGAAGGGGAAA 120

QY 121 TTAATGCACTGGAAGGAAAGGAAAGTCTTATTCCTCCAGGGAAATACGCTTAAATC 180

DB 121 TTAATGCACTGGAAGGAAAGGAAAGTCTTATTCCTCCAGGGAAATACGCTTAAATC 180

QY 181 CTAAACCTCCAGAGGATAAAGTTGCCGCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240

DB 181 CTAAACCTCCAGAGGATAAAGTTGCCGCGACGATGTTATGGGGGATATGAAGATAAGGAGC 240

QY 241 AGTTTACAGAGAGAGGGGTTATGGTTCTTTTGGCTTGGCTTACCAGATGGGATGGA 300

DB 241 AGTTTACAGAGAGAGGGGTTATGGTTCTTTTGGCTTGGCTTACCAGATGGGATGGA 300

QY 301 AACTTAACACTAATGCTCTCAATGCTCAATGAACCTCTGCAATTAAGATATGAGAG 360

DB 301 AACTTAACACTAATGCTCTCAATGCTCAATGAACCTCTGCAATTAAGATATGAGAG 360

QY 361 AGATTGTCAGATCGCATTTATTAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420

DB 361 AGATTGTCAGATCGCATTTATTAAGGCTAGAGGGTCCGGCAAGAAACCTTTACAGAGGA 420

QY 421 AACTATCAGGGGAGCAGAGGTTAGGCTTTTCAAGAGAAAGAACTCTAG 471

DB 421 AACTATCAGGGGAGCAGAGGTTAGGCTTTTCAAGAGAAAGAACTCTAG 471

RESULT 2

AAV63865

ID AAV63865 standard; DNA; 33 BP.

XX AC AAV63865;

XX AC AAV63865;

DT 01-MAR-1999 (first entry)

XX DE Polymerase enhancing factor P45 primer #1.

XX KW Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;

XX KW amplification; ligation independent cloning; primer; ss.

XX OS Synthetic.

OS *Pyrococcus furiosus*.

XX WO9842860-A1.

XX PD 01-OCT-1998.

XX PF 20-MAR-1998; 98WO-US05497.

XX PR 24-OCT-1997; 97US-0957709.

XX PR 21-MAR-1997; 97US-0822774.

XX (STRA-) STRATAGENE.

XX PI Hansen CJ, Hogrefe H;

XX WPI; 1998-542284/46.

XX PT Polymerase enhancing factor proteins, extracts and complexes -

PT improve the polymerisation activity of nucleic acid polymerases, for

PT use in amplification, sequencing and replication

XX XS Claim 84; Page 48; 161pp; English.

XX CC Oligonucleotides #1 and #2 (see AAV63866) are designed to amplify

CC a portion of a genomic clone encoding the P45 protein (see AAW72847)

CC of *Pyrococcus furiosus* DSM 3638. They include P45 and vector

CC specific sequences. Oligo #1 includes the ATG start codon and

CC 5' end of the P45 coding region (see also AAV63860). The primers

CC were designed to function with the Stratagene Affinity Protein

CC expression and purification system, which uses ligation

CC independent cloning. Recombinant P45 was expressed in *E. coli*

CC host cells. P45 is a dUTPase that enhances the activity of *P.*

CC *furiosus* DNA polymerase, thereby providing replication products

CC of greater length and purity. A claimed method of enhancing a

CC nucleic acid polymerase reaction involves adding P45 protein to

CC the reaction. A claimed DNA encoding a polymerase enhancing factor

CC (PEF) may include oligonucleotide #1, or a hybridisable or

CC degenerate variant, and a claimed method for cloning a PEF activity

CC may utilise the complement of oligonucleotide #1.

XX SQ Sequence 33 BP; 12 A; 8 C; 8 G; 5 T; 0 other;

Query Match 4.5%; Score 21; DB 19; Length 33;

Best Local Similarity 100.0%; Pred. No. 0.51;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTACTTCCAGCTGGAAT 21

DB 13 atgctacttccagactggaaa 33

RESULT 3

AAH88082

ID AAH88082 standard; cDNA; 405 BP.

XX AC AAH88082;

XX AC AAH88082;

DT 25-SEP-2001 (first entry)

XX DE Peppermint plant oil gland expressed cDNA 438.

XX KW Peppermint; plant oil gland cell; terpenoid essential oil; resin;

XX KW genetic mapping; antisense suppression; recombinant expression; ss.

XX XS Mentha x piperita.

XX PN WO200153319-A1.
 XX PD 26-JUL-2001.
 XX PF 19-JAN-2001; 2001WO-US02567.
 XX PR 20-JAN-2000; 2000US-0177264.
 XX PA (CROT/) CROTEAU R B.
 XX PA (LANG/) LANGE B M.
 XX PA (WILD/) WILDUNG M R.
 XX PI Croteau RB, Lange BM, Wildung MR;
 XX WIPI; 2001-488706/53.
 XX PT New nucleic acid molecules corresponding to mRNA molecules expressed in
 PT peppermint oil glands for enhancing expression of plant oil gland cell
 PT proteins -
 XX PS Claim 1; Page 237; 251pp; English.
 XX CC The invention relates to nucleic acid molecules (AAH87645-AAH88116) that
 CC correspond to all or part of a mRNA molecule expressed in plant oil
 CC gland cells, especially peppermint and plant oil glands that produce
 CC terpenoid essential oils and resins. The nucleic acids are useful for
 CC genetically mapping a plant genome for genes expressed in plant oil
 CC gland cells and to suppress (for example by antisense suppression) or
 CC enhance their expression (for example by genetically transforming a
 CC plant cell with a replicable expression vector that expresses one or more
 CC proteins naturally expressed in plant oil gland cells). The nucleic acids
 CC are also useful for recombinant expression of plant oil gland proteins
 CC required for terpenoid essential oil and/or resin production in bacterial
 CC and/or yeast cells.
 XX SQ Sequence 405 BP; 98 A; 84 C; 93 G; 128 T; 2 other;
 Query Match 3.8%; Score 18; DB 22; Length 405;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 349 AGATATGGAGAGAGATT 366
 Db 115 agatattggagagattt 132
 |||||
 RESULT 4
 ID AAH09366
 XX AAH09366 standard; cDNA; 571 BP.
 AC AAH09366;
 XX 26-JUN-2001 (first entry)
 XX Human cDNA clone (3'-primer) SEQ ID NO:6201.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW Homo sapiens.
 XX EP1074617-A2.
 XX 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WIPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX PS Claim 3; SEQ ID 6201; 2537pp + CD ROM; English.
 XX CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SQ Sequence 571 BP; 144 A; 162 C; 152 G; 106 T; 7 other;
 Query Match 3.8%; Score 18; DB 22; Length 571;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 GTGGAAGGAGGAGAAA 147
 Db 70 gtggaaaaggagagaaa 87
 |||||
 RESULT 5
 ID AAH88046
 XX AAH88046 standard; cDNA; 638 BP.
 AC AAH88046;
 XX 25-SEP-2001 (first entry)
 XX Peppermint plant oil gland expressed cDNA 402.
 DE Peppermint; plant oil gland cell; terpenoid essential oil; resin;
 KW genetic mapping; antisense suppression; recombinant expression; ss.
 XX Mentha x piperita.
 XX WO200153319-A1.
 XX 26-JUL-2001.
 XX 19-JAN-2001; 2001WO-US02567.
 XX 20-JAN-2000; 2000US-0177264.
 XX


```

PI Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI: 2000-579482/54.
DR P-PSDB; AAB34588.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; Page 348; 419pp; English.
PS
XX The polynucleotide sequences given in AAC59738 to AAC59787 encode the
CC human secreted proteins given in AAB34577 to AAB34626. AAB34627 to
CC AAB34686 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Example of activities include:
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC cytotatic; cardiant; vasotropic; cerebroprotective; neurotropic;
CC neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The polynucleotides and proteins can be are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also
CC used in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms and
CC cancers of the breast or liver, cardiovascular disorders,
CC cerebrovascular disorders, angiogenesis, nervous system disorders,
CC infections caused by bacteria, viruses and fungi and ocular disorders.
CC The proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues and in chemotaxis. The proteins can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59729 to AAC59737 and AAB34576 represent sequences used
CC in the exemplification of the present invention.
XX
SQ Sequence 878 BP; 192 A; 215 C; 249 G; 222 T; 0 other;

Query Match          3.8%; Score 18; DB 21; Length 878;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 GTGGAAGAAGGAGGAAAA 147
   |||||
Db 791 GTGGAAGAAGGAGGAAAA 774

RESULT 8
ID AAF33251/C
XX AAF33251 standard; cDNA; 1057 BP.
XX
XX AAF33251;
XX
XX 23-MAR-2001 (first entry)
DT
XX Human secreted protein gene 39 SEQ ID NO:49.
DE
XX
XX Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; neurotropic; anticonvulsant; vulnary;
KW antialzheimers; antiparkinsonian; antimicrobial; immune disorder;
KW multiple sclerosis; systemic lupus erythematosus; HIV; infection;
KW hyperproliferative disorder; cancer; Gaucher's disease; wound healing;
KW cardiovascular disease; Scimitar syndrome; Chaga's cardiomyopathy;
KW coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;
KW corneal graft neovascularisation; neurological disorder; regeneration;
KW Huntington's chorea; Alzheimer's disease; Parkinson's disease;
XX infectious disease; chemotaxis; ss.
XX Homo sapiens.
OS
XX

```

```

PN WO200076530-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US14933.
XX
XX 11-JUN-1999; 99US-0138572.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI: 2001-071147/08.
DR P-PSDB; AAB64920.
XX
XX Nucleic acids encoding 49 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
XX Claim 1; Page 476; 554pp; English.
XX
XX The polynucleotide sequences given in AAF33213 to AAF33261 encode the
CC human secreted proteins given in AAB64882 to AAB64930. AAB64931 to
CC AAB64991 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC vascular; antimicrobial; anti-angiogenic; ophthalmological;
CC neuroprotectant; anticonvulsant; neurotropic; antialzheimers;
CC antiparkinsonian; and vulnary. The polynucleotides and polypeptides can
CC be used in the prevention, diagnosis and treatment of diseases associated
CC with inappropriate polypeptide expression. Disorders that may be
CC prevented, diagnosed and/or treated by the above methods include immune
CC disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
CC human immuno-deficiency virus (HIV) infections), hyperproliferative
CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC neovascularisation and diabetic retinopathy), neurological disorders
CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC infectious diseases and/or for promoting wound healing, regeneration and
CC /or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 1057 BP; 238 A; 259 C; 292 G; 268 T; 0 other;

Query Match          3.8%; Score 18; DB 22; Length 1057;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 GTGGAAGAAGGAGGAAAA 147
   |||||
Db 970 GTGGAAGAAGGAGGAAAA 953

RESULT 9
ID AAT12325/C
XX AAT12325 standard; DNA; 2691 BP.
XX
XX AAT12325;
XX
XX 24-JUL-1997 (revised)
DT 29-JUN-1996 (first entry)
XX
XX DNA encoding Sulfolobus solfataricus amylase.
XX transferase; amylase; Sulfolobus; production; alpha, alpha-trehalose;
KW malto-oligosaccharide; hydrolysis; ss.
XX

```

```

OS Sulfolobus solfataricus.
XX
XX Key Location/Qualifiers
XX CDS 639..2318
XX /tag= a
XX /product= amylase
XX
XX WO9534642-A.
XX
XX PD 21-DEC-1995.
XX
XX PF 14-JUN-1995; 95WO-JP011189.
XX
XX PR 21-APR-1995; 95JP-0120673.
XX PR 15-JUN-1994; 94JP-0133354.
XX PR 18-AUG-1994; 94JP-0194223.
XX PR 31-OCT-1994; 94JP-0290394.
XX PR 21-NOV-1994; 94JP-0286917.
XX PR 21-NOV-1994; 94JP-0311185.
XX
XX (KIRI ) KIRIN BEER KK.
XX
XX WIPI; 1996-049671/05.
XX
XX P-PSDB; AAR90619.
XX
XX Sulfolobus spp. derived transferase and amylase - for production of
XX alpha, alpha-trehalose from malto-oligosaccharide(s)
XX
XX Claim 109; Page 235-240; 357pp; Japanese.
XX
XX The present sequence encodes an amylase derived from Sulfolobus
XX solfataricus. The amylase acts on a saccharide having at least
XX three sugar units which are pref glucose units at the reducing
XX end and the linkage between the first and second glucose units is
XX alpha-1,4, while the linkage between the second and third
XX glucose units is alpha-1,4 to hydrolyse alpha-1,4 linkages within
XX the sugar chain yielding alpha, alpha-trehalose and also mono- and
XX disaccharide hydrolysis products. Use of a transferase and the amylase
XX in succession on suitable substrates such as malto-oligosaccharides, is
XX useful for the production of alpha, alpha-trehalose.
XX (Revised entry submitted to correct sequence analysis breakdown.)
XX
XX Sequence 2691 BP; 925 A; 387 C; 596 G; 783 T; 0 other;
XX
XX
XX Query Match 3.8%; Score 18; DB 17; Length 2691;
XX Best Local Similarity 100.0%; Pred. No. 17;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 19 AAATCAGAAAGGAATA 36
XX |||||||||||||||
XX Db 1934 AAATCAGAAAGGAATA 1917
XX
XX RESULT 10
XX AAH17117/C
XX ID AAH17117 standard; cDNA; 3137 BP.
XX
XX AC AAH17117;
XX
XX XX 26-JUN-2001 (first entry)
XX
XX DE Human cDNA sequence SRQ ID NO:16450.
XX
XX XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX KW Homo sapiens.
XX
XX OS EP1074617-A2.
XX
XX PN 07-FEB-2001.
XX
XX PD 28-JUL-2000; 2000EP-0116126.
XX
XX PF

```

```

XX
XX 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WIPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 16450; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166, AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences, AAH03446 to
XX AAH95893 represent human amino acid sequences, and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 3137 BP; 722 A; 778 C; 865 G; 772 T; 0 other;
XX
XX
XX Query Match 3.8%; Score 18; DB 22; Length 3137;
XX Best Local Similarity 100.0%; Pred. No. 17;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 130 GTGGAAGGAGGAGGAAA 147
XX |||||||||||||||
XX Db 3068 GTGGAAGGAGGAGGAAA 3051
XX
XX RESULT 11
XX AAQ12787
XX ID AAQ12787 standard; RNA; 10723 BP.
XX
XX AC AAQ12787;
XX
XX XX 21-NOV-1991 (first entry)
XX
XX DE Dengue 2 virus genome.
XX
XX KW dengue virus; detection; consensus sequence; Flavivirus; PCR; ss.
XX
XX OS Dengue virus.
XX
XX XX Key Location/Qualifiers
XX CDS 97..10272
XX /tag= a
XX
XX FT

```

```

FT mat_peptide 712..936
FT /*tag= b
FT /product= M protein
FT mat_peptide 937..2421
FT /*tag= c
FT mat_peptide 2422..3477
FT /*tag= d
FT mat_peptide 3478..4131
FT /product= NS1
FT /*tag= e
FT mat_peptide 4132..4518
FT /product= NS2A
FT /*tag= f
FT mat_peptide 4519..6375
FT /product= NS2B
FT /*tag= g
FT mat_peptide 6376..6825
FT /product= NS3
FT /*tag= h
FT mat_peptide 6826..7569
FT /product= NS4a
FT /*tag= i
FT mat_peptide 7570..10269
FT /product= NS4B
FT /*tag= j
FT /product= NS5
XX
XX PR2654113-A.
XX
XX 10-MAY-1991.
XX
XX 09-NOV-1989; 89PR-0914724.
XX
XX 09-NOV-1989; 89PR-0014724.
XX
XX (INSP ) INST PASTEUR.
XX
XX Vincent D;
XX
XX WPI: 1991-225002/31.
XX P-PSDB; AARI3166.
XX
XX Detection and identification of Flaviviridae in biological sample
XX - by amplifying consensus sequence then hybridisation opt.
XX followed by typing, e.g. sequencing amplified prod.
XX
XX Disclosure; Fig 3; 24pp; French.
XX
XX The dengue 2 virus is an example of a member of the Flaviviridae
XX which can be identified using the probe pair of the invention. A
XX species-specific sequence can be amplified using the claimed
XX oligonucleotides as primers in a PCR reaction (see AAQ12788 and
XX AAQ12789). Other viruses which can be identified include Japanese
XX encephalitis virus and yellow fever virus.
XX
XX Sequence 10723 BP; 3557 A; 2251 C; 2647 G; 2252 U; 16 other;

Query Match 3.8%; Score 18; DB 12; Length 10723;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 14; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 445 GCGTTTTCAAAGAGAAG 462
DB 8627 gcuuuucaaagagaag 8644
||||:|||||

RESULT 12
AAH50830/c
ID AAH50830 standard; cDNA; 285 BP.
XX
XX AAH50830;

```

```

XX
XX 23-AUG-2001 (first entry)
XX
XX Human tumour associated cDNA #159.
XX
XX Human; cancer specific gene expression; gene therapy;
XX age related differential expression; ss.
XX
XX Homo sapiens.
XX
XX WO200136685-A2.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US31809.
XX
XX 17-NOV-1999; 99US-0166056.
XX 17-NOV-1999; 99US-0166106.
XX
XX (NYXI-) NYXIS NEURO THERAPIES INC.
XX
XX Kroes RA, Moskal JR, Yamamoto H;
XX
XX WPI: 2001-355647/37.
XX
XX Novel nucleic acid molecules differentially expressed in brain cancers,
XX useful for ascertaining propensity of cell for malignant phenotype or
XX ascertaining suitability of anti-neoplastic drug candidate -
XX
XX Claim 28; Page 62; 82pp; English.
XX
XX The present invention provides the sequences of 184 cDNA fragments which
XX are differentially expressed in cancer cell depending on the age of the
XX patient. They can be used to diagnose and identify treatments for
XX cancers, particularly brain cancers such as haemangioblastoma, teratoma,
XX haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The
XX present sequence is a cancer-associated cDNA of the invention.
XX
XX Sequence 285 BP; 77 A; 69 C; 45 G; 94 T; 0 other;

Query Match 3.6%; Score 17; DB 22; Length 285;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 451 TCAAAGAGAGAAAGAACT 467
DB 269 TCAAAGAGAGAAAGAACT 253
|||||

RESULT 13
AAH50736/c
ID AAH50736 standard; cDNA; 309 BP.
XX
XX AAH50736;
XX
XX 23-AUG-2001 (first entry)
XX
XX Human tumour associated cDNA #65.
XX
XX Human; cancer specific gene expression; gene therapy;
XX age related differential expression; ss.
XX
XX Homo sapiens.
XX
XX WO200136685-A2.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US31809.
XX
XX 17-NOV-1999; 99US-0166056.
XX 17-NOV-1999; 99US-0166106.
XX

```

```
XX PA (NYXI-) NYXIS NEURO THERAPIES INC.
XX PI Kroes RA, Moskal JR, Yamamoto H;
XX DR WPI: 2001-355647/37.
XX PT Novel nucleic acid molecules differentially expressed in brain cancers,
XX PT useful for ascertaining propensity of cell for malignant phenotype or
XX PT ascertaining suitability of anti-neoplastic drug candidate -
XX PS Claim 28; Page 46; 82pp; English.
XX CC The present invention provides the sequences of 184 cDNA fragments which
XX CC are differentially expressed in cancer cell depending on the age of the
XX CC patient. They can be used to diagnose and identify treatments for
XX CC cancers, particularly brain cancers such as haemangioblastoma, teratoma,
XX CC haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The
XX CC present sequence is a cancer-associated cDNA of the invention.
XX SQ Sequence 309 BP; 83 A; 75 C; 51 G; 100 T; 0 other;

Query Match 3.6%; Score 17; DB 22; Length 309;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 451 TCAAAGAGAAAGAACT 467
Db 281 TCAAAGAGAAAGAACT 265

RESULT 14
AA071175
ID AA071175 standard; DNA; 382 BP.
XX AC AA071175;
XX DT 20-MAY-1991 (first entry)
XX DE Sequence encoding IFN-alpha-2 (Arg) on plasmid parpER33.
XX KW Hybrid alpha/omega interferon; antiviral; virucide; antitumour;
XX KW cytotstatic; ss.
XX PN EP236920-A.
XX PF 04-MAR-1987; 87EP-0103030.
XX PR 10-MAR-1986; 86DE-3607835.
XX PA (BOEH ) BOEHRINGER INGELHEIM.
XX PI Hauptmann R, Swetly P, Meindl P, Gunther A, Falkner E;
XX PI Bodo G, Maurer-Fogy I;
XX DR WPI: 1987-258223/37.
XX DE New hybrid interferon prods. useful as antiviral agents - contg.
XX PT alpha and omega interferon fragments
XX PS Disclosure; p9; 65pp; German.
XX CC Hybrid interferons comprising a fragment of an alpha-interferon and
XX CC a fragment of an omega-interferon, and their N-terminal Met or
XX CC N-formyl-Met derivs. and N-glycosylated derivs., are new. The hybrid
XX CC interferons are useful as antiviral and antitumour agents.
XX SQ Sequence 382 BP; 111 A; 91 C; 99 G; 80 T; 1 other;

Query Match 3.6%; Score 17; DB 8; Length 382;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 CAGACTGGAAATCAGA 27
Db 102 cagactggaaatcaga 118

RESULT 15
AA070530
ID AA070530 standard; DNA; 383 BP.
XX AC AA070530;
XX DT 20-MAY-1991 (first entry)
XX DE Sequence encoding IFN-alpha-2 (Arg) on plasmid parpER33.
XX KW Hybrid alpha/omega interferon; antiviral; virucide; antitumour;
XX KW cytotstatic; ss.
XX PN EP236920-A.
XX PF 04-MAR-1987; 87EP-0103030.
XX PR 10-MAR-1986; 86DE-3607835.
XX PA (BOEH ) BOEHRINGER INGELHEIM.
XX PI Hauptmann R, Swetly P, Meindl P, Gunther A, Falkner E;
XX PI Bodo G, Maurer-Fogy I;
XX DR WPI: 1987-258223/37.
XX DE New hybrid interferon prods. useful as antiviral agents - contg.
XX PT alpha and omega interferon fragments
XX PS Disclosure; p9; 65pp; German.
XX CC Hybrid interferons comprising a fragment of an alpha-interferon and
XX CC a fragment of an omega-interferon, and their N-terminal Met or
XX CC N-formyl-Met derivs. and N-glycosylated derivs., are new. The hybrid
XX CC interferons are useful as antiviral and antitumour agents.
XX SQ Sequence 383 BP; 111 A; 91 C; 100 G; 80 T; 1 other;

Query Match 3.6%; Score 17; DB 8; Length 383;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 CAGACTGGAAATCAGA 27
Db 102 cagactggaaatcaga 118

Search completed: January 31, 2002, 12:17:54
Job time: 126 sec
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 11:49:27 ; Search time 1331.66 Seconds
(without alignments)
5834.952 Million cell updates/sec

Title: US-08-957-709-70

Perfect score: 471

Sequence: 1 ATGCTACTTCCAGACTGGAA.....CAAGAGAGAAAGAACTCTAG 471

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.om.*

20: em.or.*

21: em.ov.*

22: em.pat.*

23: em.ph.*

24: em.pl.*

25: em.ro.*

26: em.sts.*

27: em.sy.*

28: em.un.*

29: em.vi.*

30: em.htgo_hum.*

31: em.htgo_inv.*

32: em.htgo_rod.*

33: em.htg_hum.*

34: em.htg_inv.*

35: em.htg_rod.*

36: em.htg_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

*

Result No.	Score	Query Match	Length	DB	ID	Description
1	23	4.9	253505	1	AP000007	AP000007 Pyrococcus
2	21	4.5	59777	3	AC084454	AC084454 Caenorhab
3	21	4.5	102195	2	AC091338	AC091338 Rattus no
4	21	4.5	162041	2	AC025947	AC025947 Homo sapi
5	20	4.2	1686	14	AF331362	AF331362 HTV1 isol
6	20	4.2	93859	8	AC021666	AC021666 Arabidops
7	20	4.2	104806	8	AC006228	AC006228 Genomic s
8	20	4.2	146393	2	AC012124	AC012124 Homo sapi
9	20	4.2	164935	9	CNSO1DUN	AL133299 Human chr
10	20	4.2	175291	2	AC024429	AC024429 Homo sapi
11	20	4.2	187479	2	AC093290	AC093290 Homo sapi
12	20	4.2	190174	2	AC022739	AC022739 Homo sapi
13	20	4.2	192725	2	AC087841	AC087841 Mus muscu
14	19	4.0	380	14	HIVU69075	U69075 HIV-1 patie
15	19	4.0	418	11	HSPE48B07	AL033810 H.sapiens
16	19	4.0	984	14	AY000296	AY000296 HIV-1 iso
17	19	4.0	984	14	AY000384	AY000384 HIV-1 iso
18	19	4.0	984	14	AY000894	AY000894 HIV-1 iso
19	19	4.0	984	14	AY002306	AY002306 HIV-1 iso
20	19	4.0	1680	14	AF331408	AF331408 HTV1 isol
21	19	4.0	1686	14	AF331387	AF331387 HTV1 isol
22	19	4.0	32785	2	AC013560	AC013560 Homo sapi
23	19	4.0	34576	3	CET23D8	281128 Caenorhabdi
24	19	4.0	35012	2	AF314059	AF314059 Homo sapi
25	19	4.0	47300	9	HSMTM1	Y15994 Homo sapien
26	19	4.0	61502	2	AC015619	AC015619 Homo sapi
27	19	4.0	105443	9	AL160256	AL160256 Human DNA
28	19	4.0	109442	9	AF238378	AF238378 Homo sapi
29	19	4.0	111680	2	AF200455	AF200455 Homo sapi
30	19	4.0	113026	2	AL354748	AL354748 Homo sapi
31	19	4.0	114413	2	AC092286	AC092286 Homo sapi
32	19	4.0	117993	2	AC091284	AC091284 Mus muscu
33	19	4.0	123585	9	AC004707	AC004707 Homo sapi
34	19	4.0	129090	2	AC083961	AC083961 Homo sapi
35	19	4.0	129402	2	AP002825	AP002825 Homo sapi
36	19	4.0	133072	9	HS1097P24	AL035258 Human DNA
37	19	4.0	143092	3	CEY17G7B	AL023828 Caenorhab
38	19	4.0	143738	9	AC006054	AC006054 Homo sapi
39	19	4.0	144079	2	AC024342	AC024342 Homo sapi
40	19	4.0	146810	9	AL159170	AL159170 Human DNA
41	19	4.0	150351	2	AP002505	AP002505 Homo sapi
42	19	4.0	150478	2	AF233439	AF233439 Homo sapi
43	19	4.0	156955	2	AC009419	AC009419 Homo sapi
44	19	4.0	157579	9	AC018639	AC018639 Human chr
45	19	4.0	163136	9	HS15D23	AL031599 Human DNA

ALIGNMENTS

RESULT 1

AP000007

LOCUS

DEFINITION

Pyrococcus horikoshii OT3 genomic DNA, 1485001-1738505 nt. position (777).

ACCESSION

AP000007 AB009464 AB009521 AB009522 AB009523 AB009524

AB009525 AB009526 AB009527 AB009528 AB009529 AB009530 AB009531

AP000007.1 GI:3236134

VERSION

KEYWORDS

SOURCE

ORGANISM

Pyrococcus horikoshii (strain:OT3) DNA.

Pyrococcus horikoshii

Archaea; Euryarchaeota; Thermococcales; Thermococcaceae;

Pyrococcus.

REFERENCE

1 (bases 1 to 253505)

Tanaka,T., Kawarabayasi,Y. and Kikuchi,H.

Direct Submission

Submitted (11-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yutaka

Kawarabayasi, National Institute of Technology and Evaluation,

Biotechnology Center, 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo

151-0066, Japan (E-mail:genomeOT3@nitech.go.jp, Tel:++81-3-3481-8951,

/translation="MFILIVVRPGLNEVYKLDKRPVEKYRILYRGDYDIIRLSVE
 GKALDEFHDKVLKRLGRIEISSETLIASSY"
 gene complement(3398. .4735)
 CDS /gene="PH1694"
 complement(3398. .4735)
 /gene="PH1694"
 /note="similar to PIR:B64312 percent identity:60.870 in
 444aa; owl:AFFH1 percent identity:50.236 in 430aa;
 owl:SAFFH percent identity:50.611 in 415aa.
 motif-ATP/GTP-binding site motif A (P-loop) ; SRP54-type
 proteins GTP-binding domain signature"
 /codon_start=1
 /transl_table=11
 /product="445aa long hypothetical signal recognition
 particle protein"
 /protein_id="BAA30807.1"
 /db_xref="GI:3258124"
 /translation="MWMVLDLGRALSNALKKIARAGSVDEALVKEVVRDIQRLALQ
 DVNRLVRLKTEIQRRALKEPPAGISKEHLIKIVYBELTKFLGTEAKPIEIKKP
 TVLLTVGVGSGKTTTVAKLARYFORGYKGVVCSDTWRPGAYHQLKLLDPYHIEV
 FGDPEKDAVKLAKEGVFKTRVDLIIVDTAGRHKEKDLIEEMRMISSEIRPHEV
 ILVIDGTIGQAYNOALAFKEATPIGSIIVTKLDSAKGGGALSAVAATGAKPIKFGV
 GEKIDLEFPDARFVSRLLGLDIOGLEKFEKEFEVEEDIDRELGRKFTLKD
 MTAQLGARMRGPKGLIKRMIPGLISLDELISVGEERLRKFKVIMDSRTEELNRP
 EIINYSRIKRIARAGSITKDKVKELLTYNQMKKFKSMNKRQLSLARFGM"
 complement(4767. .5384)
 /gene="PH1695"
 complement(4767. .5384)
 /gene="PH1695"
 /note="similar to PIR:F64336 percent identity: 42.778 in
 196aa. motif-ATP/GTP-binding site motif A (P-loop)"
 /codon_start=1
 /transl_table=11
 /product="205aa long hypothetical thymidylate kinase"
 /protein_id="BAA30808.1"
 /db_xref="GI:3258125"
 Query Match 4.9%; Score 23; DB 1; Length 253505;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 151 GTATTCTCTCAAGGGGAATACGC 173
 Db 251005 GTATTCTCTCAAGGGGAATACGC 251027
 RESULT 2
 AC084454
 LOCUS
 DEFINITION
 AC084454
 VERSION
 AC084454.1 GI:11094904
 KEYWORDS
 HTG.
 SOURCE
 Caenorhabditis briggsae.
 Caenorhabditis briggsae.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 59777)
 Washington University Genome Sequencing Center.
 The C. briggsae Genome Sequencing Project
 JOURNAL
 Unpublished
 REFERENCE
 2 (bases 1 to 59777)
 WATERSTON, R.
 Direct Submission
 Submitted (04-NOV-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA
 e-mail: jspieth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.
 FEATURES
 source
 1. 59777
 Location/Qualifiers
 /organism="Caenorhabditis briggsae"
 /strain="GutAraT G16"
 /db_xref="taxon:6238"
 /clone="CB027J01"
 BASE COUNT 19316 a 10945 c 11417 g 18099 t
 ORIGIN
 Query Match 4.5%; Score 21; DB 3; Length 59777;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 49 TTTTCTGAAGAATCGTCCAA 69
 Db 47640 TTTTCTGAAGAATCGTCCAA 47660
 RESULT 3
 AC091338
 LOCUS
 DEFINITION
 AC091338
 VERSION
 AC091338.2 GI:13702744
 KEYWORDS
 HTG; HTGS_PHASE1.
 SOURCE
 Norway rat.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 102195)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
 Alsbrooks, S.L., Anaratunge, H.C., Are, J.R., Banks, T., Barbara, J.,
 Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
 Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C.,
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carlton, T.F.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Haves, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C.,
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Louissegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
 Ogum, M., Okwunonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perwez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, K., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,

Thu Jan 31 13:32:20 2002

us-08-957-709-70.rge

Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 102195)

Worley, K. C.
Direct Submission
Submitted (18-Apr-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 21, 2001 this sequence version replaced gi:13661904.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Center project name: TUCG

Center clone name: CH230-1C12

----- Summary Statistics -----

Sequencing vector: Plasmid; M77789

Chemistry: Dye-Primer Body; 65% of reads

Chemistry: Dye-terminator Big Dye; 35% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 99820 bases at least Q40

Consensus quality: 116847 bases at least Q30

Consensus quality: 125383 bases at least Q20

Estimated insert size: 90042; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 55 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 3199: contig of 3199 bp in length

* 3200: gap of unknown length

* 3201: contig of 2566 bp in length

* 3202: gap of unknown length

* 3203: contig of 2677 bp in length

* 3204: gap of unknown length

* 3205: contig of 2723 bp in length

* 3206: gap of unknown length

* 3207: contig of 2248 bp in length

* 3208: gap of unknown length

* 3209: contig of 2583 bp in length

* 3210: gap of unknown length

* 3211: contig of 2075 bp in length

* 3212: gap of unknown length

* 3213: contig of 2320 bp in length

* 3214: gap of unknown length

* 3215: contig of 2432 bp in length

* 3216: gap of unknown length

* 3217: contig of 1734 bp in length

* 3218: gap of unknown length

* 3219: contig of 2905 bp in length

* 3220: gap of unknown length

* 3221: contig of 2661 bp in length

* 3222: gap of unknown length

* 3223: contig of 2238 bp in length

* 3224: gap of unknown length

* 3225: contig of 2338 bp in length

* 3226: gap of unknown length

* 3227: contig of 1896 bp in length

* 3228: gap of unknown length

39016: contig of 1976 bp in length
40991: gap of unknown length
41091: contig of 1823 bp in length
42914: gap of unknown length
43014: contig of 2513 bp in length
43015: gap of unknown length
45527: contig of 1555 bp in length
45628: gap of unknown length
47182: contig of 1197 bp in length
47282: gap of unknown length
48479: contig of 2768 bp in length
48579: gap of unknown length
48580: contig of 1384 bp in length
51348: gap of unknown length
52831: contig of 1269 bp in length
52832: gap of unknown length
52932: contig of 2114 bp in length
54301: gap of unknown length
56144: contig of 1274 bp in length
56515: gap of unknown length
57888: contig of 1432 bp in length
59320: gap of unknown length
59420: contig of 1897 bp in length
60528: gap of unknown length
60529: contig of 1366 bp in length
62855: gap of unknown length
64021: contig of 1546 bp in length
64121: gap of unknown length
65667: contig of 1837 bp in length
65668: gap of unknown length
67604: contig of 1536 bp in length
67704: gap of unknown length
69240: contig of 1307 bp in length
69341: gap of unknown length
70647: contig of 1080 bp in length
70748: gap of unknown length
71828: contig of 1399 bp in length
71937: gap of unknown length
73326: contig of 1421 bp in length
73426: gap of unknown length
73327: contig of 1421 bp in length
74847: gap of unknown length
74947: contig of 1421 bp in length
74948: gap of unknown length
76369: contig of 1234 bp in length
76469: gap of unknown length
77702: contig of 1718 bp in length
77802: gap of unknown length
79520: contig of 1875 bp in length
79621: gap of unknown length
81496: contig of 1400 bp in length
81596: gap of unknown length
82995: contig of 1148 bp in length
83095: gap of unknown length
84243: contig of 1506 bp in length
84343: gap of unknown length
84344: contig of 1132 bp in length
85850: gap of unknown length
85949: contig of 1349 bp in length
87082: gap of unknown length
87181: contig of 1256 bp in length
88630: gap of unknown length
88631: contig of 1238 bp in length
89887: gap of unknown length
91224: contig of 1107 bp in length
91225: gap of unknown length
91325: contig of 1317 bp in length
92432: gap of unknown length
92532: contig of 1323 bp in length
93849: gap of unknown length
93949: contig of 1353 bp in length
95272: gap of unknown length
95273: contig of 1468 bp in length
96725: gap of unknown length
96824: contig of 1468 bp in length
98292: gap of unknown length

```

* 98293 98392: gap of unknown length
* 98393 99454: contig of 1062 bp in length
* 99455 99554: gap of unknown length
* 100912: contig of 1358 bp in length
* 100913 101012: gap of unknown length
* 101013 102195: contig of 1183 bp in length.

Query Match      4.5%; Score 21; DB 2; Length 102195;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 AGAACCCCTTACAGAGAAAC 423
Db 47512 AGAACCCCTTACAGAGAAAC 47532

RESULT 4
AC025947
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-78A18, WORKING DRAFT
SEQUENCE 6 unordered pieces.
AC025947
AC025947.8 GI:14196356
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162041)
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 162041)
Smith,D.R.
Direct Submission
Submitted (18-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On May 25, 2001 this sequence version replaced gi:14190658.
----- Genome Center
Center: Genome Therapeutics Corporation
Center code: GTC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
----- Project Information
Center project name: hg307
----- Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 158813 bases at least Q40
Consensus quality: 159433 bases at least Q30
Consensus quality: 159760 bases at least Q20
Insert size: 161640; sum-of-contigs
Quality coverage: 7.4x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1574: contig of 1574 bp in length
* 1575 1675: gap of unknown length
* 1675 12928: contig of 11254 bp in length
* 12928 13029: gap of unknown length
* 13029 28414: contig of 15386 bp in length
* 28414 28515: gap of unknown length
* 28515 53750: contig of 25236 bp in length
* 53750 53851: gap of unknown length
* 53851 95649: contig of 41799 bp in length

```

```

* 95650 95749: gap of unknown length
* 95750 162041: contig of 66292 bp in length.

FEATURES
Source
1..162041
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RPCI-11"
/clone="RP11-78A18"
/misc_feature 1..1574
/note="assembly_name:Contig20"
misc_feature 1675..12928
/note="assembly_name:Contig24
clone_end:SP6"
misc_feature 13029..28414
/note="assembly_name:Contig25"
misc_feature 28515..53750
/note="assembly_name:Contig26"
misc_feature 53851..95649
/note="assembly_name:Contig27"
misc_feature 95750..162041
/note="assembly_name:Contig28
clone_end:T7"

BASE COUNT 48263 a 34100 c 33559 g 45618 t 501 others
ORIGIN

Query Match      4.5%; Score 21; DB 2; Length 162041;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 AAAATCAGAAAAGAAATACTT 39
Db 144239 AAAATCAGAAAAGAAATACTT 144259

RESULT 5
AF331362
LOCUS
DEFINITION HIV1 isolate T9C16 from the Netherlands pol polyprotein (pol) gene,
partial cds.
ACCESSION AF331362
VERSION AF331362.1 GI:13095339
KEYWORDS
SOURCE Human immunodeficiency virus type 1.
ORGANISM Human immunodeficiency virus type 1.
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate
lentivirus group.
REFERENCE 1 (bases 1 to 1686)
AUTHORS Lukashov,V.V., Huismans,R., Jebbink,M.F., Danner,S.A., de Boer,R.J.
and Goudsmit,J.
TITLE Selection by AZT and rapid replacement in the absence of drugs of
HIV type 1 resistant to multiple nucleoside analogs
JOURNAL AIDS Res. Hum. Retroviruses (2001) In press
REFERENCE 2 (bases 1 to 1686)
AUTHORS Lukashov,V.V.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-2000) Human Retrovirology, Academic Medical
Center, University of Amsterdam, Meibergdreef 15, Amsterdam 1105
Az, The Netherlands
FEATURES
Source
1..1686
/organism="Human immunodeficiency virus type 1"
/virion
/isolate="T9C16"
/db_xref="taxon:11676"
/country="Netherlands"
/note="This sequence is from a set of 231 sequences of
HIV-1 RTs obtained from a single individual at 12 time
points over 6 years. In the isolate names, T shows time
point, C shows clone number."
<1..>1686
/gene="pol"

```

```
CDS
<1. >1686
/gene="pol"
/codon_start=1
/product="pol polyprotein"
/protein_id="AA12234.1"
/db_xref="GI:13095340"
/translation="PISPIETVPLKPGMDGPKVKQWPLTEEKIKALVEITEMEKE
GKVSITGPNPNYNTVFAIKSSSRGTRKLVDFRELNRKQDFWEVLGPHAGL
KKKSTVLVDGDAFVSLDKDFRYTAFTIPSTNNETPGIRYNLVPQWKGSPA
IFQCSMTKLDVFRKQNPDIYQYMDLYVGSDEIEGRRIKEELRHLRWGPT
PKKHQEPFLWGLYELHPDKWTQPIQLPEKDSWTVNDLQKLVGLKNAWQIYAGI
KVRLQKLLRGAKSLETVVQLTEAELEAENREILKEPVGHVYDPSKDLAEVQKQ
GQOMTYIOEPFNKLTGKYAKMKGAHTNDVKLTAEVQKIARESIVIGKTPFKF
LPIQKETAETVQKQVYSITDTQKTELQAIHLAQDSLEVNIVTDSYALGII
ETKLKAGVTVQKQVYSITDTQKTELQAIHLAQDSLEVNIVTDSYALGII
QAPQDSSELSVQIEQLIKKEKVVYLAWVPAHKGGGQNEQVQDKLVASGIRKVL"
BASE COUNT 672 a 273 c 372 g 369 t
ORIGIN

Query Match 4.2%; Score 20; DB 14; Length 1686;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAAGGAAGGAAAAGTC 150
|||||
Db 122 TGGAAAAGGAAGGAAAAGTC 141

RESULT 6
AC021666 93859 bp DNA PLN 19-JAN-2001
Arabidopsis thaliana chromosome 1 BAC T22A15 genomic sequence,
complete sequence.
AC021666
AC021666.5 GI:12324708
HFG.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 93859)
Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Renning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC T22A15 genomic sequence
Unpublished
2 (bases 1 to 93859)
Lin, X. and Kaul, S.
Direct Submission
Submitted (19-JAN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 93859)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280867.
Address all correspondence to: at@tigr.org

BAC clone T22A15 is from Arabidopsis thaliana chromosome 1
clone.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html) GenemarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/) Glimmer (a variant
of Glimmer, see Michael Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
Gensplitter (Mihaila Pertea and Steven Salzberg, contact
mperleat@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR

(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers
1. 93859
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="T22A15"
/complement(633..702)
/rpt_family="AT_rich"
/complement(819..848)
/rpt_family="AT_rich"
/complement(1021..1129)
/rpt_family="novel repeat related to GB|AB000094 inorganic
phosphate transporter pseudogene. Matches fragments of the
ARNOLDY non-autonomous DNA-transposon family."
/complement(1093..1184)
/rpt_family="AT_rich"
/complement(1362..3403)
/rpt_family="AT_rich"
Join(<4584..4631,4701..4874,4967..5056,6271..6390,
6492..6602,6702..6793,6901..7006,7106..7192,7393..7806)
/gene="T22A15.1"
4584..7806
/note="Similar to hypothetical protein GB|AAF18647"
Join(4584..4631,4701..4874,4967..5056,6271..6390,
6492..6602,6702..6793,6901..7006,7106..7192,7393..7806)
/gene="T22A15.1"
/codon_start=1
/product="unknown protein; 4584-7806"
/protein_id="AAG52318.1"
/db_xref="GI:12324715"
/translation="MEQTKDSDFFEDCKKMETDTSQAEARNKDIQGIOWEGPKY
NRDEFRETRKQYDNFNITLRFSSREKSLQKHEGVVKGKNFYDFQFNRLVTSIVH
FQLRLNVAWSKHVYLMQYSLMWSLLQKQEVINVARSKLPLFSEPLSRQVQS
SMAVENLILGDFGELCKVCNQGFACFTRLSCTEDNATNTVDIYDRDPSGLRLI
TANNCKIRVDAQSFTRVSEFALDMSVNTSVSPDKLLAVVQDTECLISDSHSGK
VISSLRGHRDYSFASAWHPNGLILATNQDTACRLDIRNPSEFAVLKGNMGAIRGL
KETPEGRFLAAEPADVFHFDFTQSGFLQSQEIDLFGELTAGISFSPDTEALYVGVADR
TYGSLLEYKRRKDNHYMDSFY"
8286..8340
/rpt_family="(CCCTAA)n"
/complement(8520..8557)
/rpt_family="(TA)n"
/complement(8625..8667)
/rpt_family="(TA)n"
/complement(9197..9287)
/rpt_family="AT_rich"
/complement(9378..9400)
/rpt_family="AT_rich"
/complement(9490..9569)
/rpt_family="AT_rich"
/complement(10188..10218)
/rpt_family="AT_rich"
10951..11016
/rpt_family="(TGG)n"
/complement(11112..11139)
/rpt_family="AT_rich"
11416..11462 (TAAAA)n"
/rpt_family="AT_rich"
/complement(11489..11518)
/rpt_family="AT_rich"
```

```
complement(12256..12293)
/rpt_family="AT_rich"
repeat_region
complement(13262..13309)
/rpt_family="AT_rich"
repeat_region
complement(13380..13407)
/rpt_family="AT_rich"
repeat_region
complement(15579..15599)
/rpt_family="AT_rich"
repeat_region
complement(15627..15663)
/rpt_family="AT_rich"
repeat_region
complement(15698..15764)
/rpt_family="AT_rich"
repeat_region
complement(15804..15825)
/rpt_family="AT_rich"
repeat_region
complement(16022..16070)
/rpt_family="AT_rich"
repeat_region
complement(16458..16494)
/rpt_family="AT_rich"
repeat_region
complement(17209..17261)
/rpt_family="AT_rich"
repeat_region
complement(18187..18233)
/rpt_family="AT_rich"
repeat_region
complement(18417..18460)
/rpt_family="AT_rich"
repeat_region
complement(19233..19258)
/rpt_family="AT_rich"
repeat_region
19272..>20248
/genes="T22A15.2"
mRNA
19272..20248
/genes="T22A15.2"
gene
19272..20248
/genes="T22A15.2"
CDS
19304..20248
/notes="similar to AP2 domain containing protein RAP2.4,
GB:AA049770 from [Arabidopsis thaliana]"
/genes="T22A15.2"
/codon_start=1
/product="putative AP2 domain-containing transcription
factor; 19304-20248"
/protein_id="AAG52316.1"
/db_xref="GI:12324713"
/translation="MADLFGGSGGELMEALQPFYKASSTASNPASFASNDASFAP
NLFSSSYNPHASLPFSSHTSYDIIYSGSWTSPSSGSLQDQPPHQFQSHYQNT
IYTHODNCTMLFTEPSPGFMTQPPSGSVSKPAKLYRGVRRHGGKVAEIRL
PNRRLTGLTDEAEALAYDRAAFKLRGSAALNFPALRYQTSSPSDTEGYPI
QAAVDAKRAILAEKPAKQKTERTSKRKAASAEQPSAPQHSQSGSDGSGSP
TSDVMVQECQPEMWNFNMLGKPCSYEIDWASTLS"
complement(join(<1797..22002,22089..22155,22238..22330,
22416..22499,22562..22613,22690..22757,22844..23023,
23276..23368,23612..23692,23857..23937,24032..24115,
24199..24272,24358..>24499))
/genes="T22A15.3"
gene
complement(21797..24499)
/genes="T22A15.3"
CDS
complement(join(21911..22002,22089..22155,22238..22330,
22416..22499,22562..22613,22690..22757,22844..23023,
23276..23368,23612..23692,23857..23937,24032..24115,
24199..24272,24358..>24499))
/genes="T22A15.3"
/codon_start=1
/product="unknown protein; 24499-21911"
/protein_id="AAG52317.1"
/db_xref="GI:12324714"
/translation="MAGILNKLRLDAPKINEDFYRSLTSGGVITLLSSVNFLLFF
SELRLYLHVTFKLVDPTRGETLRINFDIFPALACSILSDVAMDISGELHLDVKH
DIKRLDSNGVTIERQDGATKLENLQKHGGRIHNETYCGSIGAEAEHDC
NSCEDVREAYRRKGVGNPDLDIDQKREGFLQVRKDEEGCNIGFLEVNKVGFP
HPAGKSFQSGVHDLAFQDSFNISHLNRLTYGDYFFGVNPLDKVMSQDTP
NAMYQVYFVTVTDIRGHTIQSNQSVTEHVRKSEAGQLSPGVFFVDLSPIK
VITFEHISFLHETNVCAIGVFTVSIGIIDAFIVHGOKAIKKKEICKFS"
complement(24685..24714)
/rpt_family="(GAA)n"
repeat_region
complement(25198..25397)
/rpt_family="SIMPLEGUY1|SIMPLEGUY1 SIMPLEGUY1 is a
putative non-autonomous DNA transposon - a consensus."
```

```
complement(25564..25619)
/rpt_family="AT_rich"
repeat_region
complement(26152..26199)
/rpt_family="AT_rich"
repeat_region
complement(29580..29665)
/rpt_family="(A)n"
gene
complement(29824..34890)
/genes="T22A15.4"
/notes="pseudogene, putative gag-pol polyprotein"
/pseudo
repeat_region
complement(30758..31044)
/rpt_family="TORTL1|TORTL1 Lycopersicon esculentum"
Query Match 4.2%; Score 20; DB 8; Length 93859;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 448 TTTTCAAGAGAGAAAGAACT 467
|||||
Db 3660 TTTTCAAGAGAGAAAGAACT 3679

RESULT 7
AC006228/c AC006228 104806 bp DNA PLN 05-JAN-2001
LOCUS Genomic sequence for Arabidopsis thaliana BAC F5J5, complete
DEFINITION sequence.
AC006228 AC006228.5 GI:12039051
VERSION HTG.
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 104806)
Chao,Q., Shinn,P., Dunn,P., Buehler,E., Kahn,S., Kim,C., Walker,M.,
Williams,S., Altafi,H., Araujo,R., Conn,L., Conway,A.B.,
Gonzalez,A., Hansen,N.F., Huizar,L., Kremenetskaia,I., Lenz,C.,
Li,J., Liu,S., Luros,S., Rowley,D., Schwartz,J., Toriumi,M.,
Vysotskaia,V., Yu,G., Davis,R.W., Federspiel,N.A., Theologis,A. and
Ecker,J.R.
Genomic sequence for Arabidopsis thaliana BAC F5J5
Unpublished
REFERENCE 2 (bases 1 to 104806)
Ecker,J.R.
Direct Submission
Submitted (18-DEC-1998) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 104806)
Ecker,J.R.
Direct Submission
Submitted (11-APR-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
4 (bases 1 to 104806)
Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
Shinn,P., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L.,
Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,
Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
Thavari,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
Federspiel,N., Theologis,A. and Ecker,J.
Direct Submission
Submitted (17-DEC-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
5 (bases 1 to 104806)
Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
Shinn,P., Altafi,H., Bei,B., Chin,C., Chlou,J., Choi,E., Conn,L.,
Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,
Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
Thavari,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
Federspiel,N., Theologis,A. and Ecker,J.
Direct Submission
Submitted (17-DEC-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
```


/evidence-not_experimental
 /product="F5J5.7"
 /protein_id="AAF18646.1"
 /db_xref="GI:6598591"
 /translation="MRMKLRISKKIRIMCVNMRKSRFTOSFIPLVGLDSVCFH"

Query Match 4.2%; Score 20; DB 8; Length 104806;
 Best Local Similarity 100.0%; Pred. No. 6.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 448 TTTTCAAGAGAAGAACT 467
 |||||
 Db 32053 TTTTCAAGAGAAGAACT 32034

RESULT 8
 AC012124/c
 LOCUS AC012124 146393 bp DNA HTG 12-MAR-2000
 DEFINITION Homo sapiens clone RP11-3B7, WORKING DRAFT SEQUENCE, 23 unordered pieces.

AC012124
 VERSION AC012124.4 GI:7107859
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 146393)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 1 (bases 1 to 146393)
 TITLE Homo sapiens, clone RP11-3B7
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 146393)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy L., Boukhaalter,B.,
 Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeRellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald P., Marquis N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Feb 28, 2000 this sequence version replaced gi:6479073.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L2707
 Center clone name: 3-B-7

----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 110026 bases at least Q40
 Consensus quality: 130435 bases at least Q30
 Consensus quality: 138989 bases at least Q20
 Insert size: 165000; agarose-fp
 Insert size: 144193; sum-of-contigs
 Quality coverage: 4.9 in Q20 bases; agarose-fp
 Quality coverage: 5.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1110: contig of 1110 bp in length
 * 1111 1210: gap of 100 bp
 * 1211 2297: contig of 1087 bp in length
 * 2298 2397: gap of 100 bp
 * 2398 4438: contig of 2041 bp in length
 * 4439 4538: gap of 100 bp
 * 4539 6653: contig of 2115 bp in length
 * 6654 6753: gap of 100 bp
 * 6754 8250: contig of 1497 bp in length
 * 8251 8350: gap of 100 bp
 * 8351 11796: contig of 3446 bp in length
 * 11797 11896: gap of 100 bp
 * 11897 14787: contig of 2891 bp in length
 * 14788 14887: gap of 100 bp
 * 14888 18648: contig of 3761 bp in length
 * 18649 18748: gap of 100 bp
 * 18749 23568: contig of 4820 bp in length
 * 23569 23668: gap of 100 bp
 * 23669 28622: contig of 4954 bp in length
 * 28623 28722: gap of 100 bp
 * 28723 32571: contig of 3849 bp in length
 * 32572 32671: gap of 100 bp
 * 32672 37334: contig of 4663 bp in length
 * 37335 37434: gap of 100 bp
 * 37435 44015: contig of 6581 bp in length
 * 44016 44115: gap of 100 bp
 * 44116 50291: contig of 6176 bp in length
 * 50292 50391: gap of 100 bp
 * 50392 56846: contig of 6455 bp in length
 * 56847 56946: gap of 100 bp
 * 56947 65088: contig of 8142 bp in length
 * 65089 65188: gap of 100 bp
 * 65189 69826: contig of 4638 bp in length
 * 69827 69926: gap of 100 bp
 * 69927 80045: contig of 10119 bp in length
 * 80046 80145: gap of 100 bp
 * 80146 88803: contig of 8658 bp in length
 * 88804 88903: gap of 100 bp
 * 88904 98614: contig of 9711 bp in length
 * 98615 98714: gap of 100 bp
 * 98715 107307: contig of 8593 bp in length
 * 107308 107407: gap of 100 bp
 * 107408 120788: contig of 13381 bp in length
 * 120789 120888: gap of 100 bp
 * 120889 146393: contig of 25505 bp in length.

FEATURES

source
 1..146393
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RP11-3B7"
 /clone_lib="RP11-11 Human Male BAC"
 1..1110
 /note="assembly_fragment"
 1211..2297
 /note="assembly_fragment"
 2398..4438
 /note="assembly_fragment"
 4539..6653
 /note="assembly_fragment"
 6754..8250
 /note="assembly_fragment"
 8351..11796
 /note="assembly_fragment"
 11897..14787
 /note="assembly_fragment"
 misc_feature
 1211..2297
 misc_feature
 2398..4438
 misc_feature
 4539..6653
 misc_feature
 6754..8250
 misc_feature
 8351..11796
 misc_feature
 11897..14787

```
misc_feature 14888..18648
/note="assembly_fragment"
misc_feature 18749..23568
/note="assembly_fragment"
misc_feature 23669..28622
/note="assembly_fragment"
misc_feature 28723..32571
/note="assembly_fragment"
misc_feature 32672..37334
/note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 37435..44015
/note="assembly_fragment"
misc_feature 44116..50291
/note="assembly_fragment"
misc_feature 50392..56846
/note="assembly_fragment"
misc_feature 56947..63088
/note="assembly_fragment"
misc_feature 63189..69826
/note="assembly_fragment"
misc_feature 69927..80045
/note="assembly_fragment"
misc_feature 80146..88803
/note="assembly_fragment"
misc_feature 88904..98614
/note="assembly_fragment"
misc_feature 98715..107307
/note="assembly_fragment"
misc_feature 107408..120788
/note="assembly_fragment"
misc_feature 120889..146393
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
BASE COUNT 38984 a 34000 c 33760 g 37440 t 2209 others
ORIGIN

Query Match 4.2%: Score 20; DB 2; Length 146393;
Best Local Similarity 100.0%; Pred No. 6.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 AACATATGGAGAGAGATTG 367
|||||
DB 99355 AACATATGGAGAGAGATTG 99336

RESULT 9
CNSOLIDUN/c CNSOLIDUN 164935 bp DNA PRI 23-MAY-2001
LOCUS human chromosome 14 DNA sequence BAC R-62H20 of library RPCI-11
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL133299
VERSION AL133299.4 GI:14268332
KEYWORDS HTG; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164935)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissensbach,J.
Sequencing of the human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 164935)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT
- Web : www.genoscope.cns.fr)
On May 30, 2001 this sequence version replaced gi:12140269.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-2190G12
Downstream BAC (overlapping the SP6 end) : R-16B13 (AC-AL157911)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 9.39% in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 - 9 : 4
10 - 19 : 73
20 - 29 : 134
30 - 39 : 298
40 - 49 : 1615
50 - 59 : 5427
60 - 69 : 9825
70 - 79 : 20561
80 - 89 : 44266
90 - 99 : 82732
-----
Percentage of bases with a quality value >= 40 : 99 %
-----
FEATURES
source
1..164935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="RPCI-11"
122807..122993
/note="matching EMBL:G11072
RHdb:RH33369
dbSTS:STS31929
Identified using the e-PCR software (G. Schuler)"
136777..136955
/note="matching EMBL:AI033968
RHdb:RH103549
dbSTS:STS71078
Identified using the e-PCR software (G. Schuler)"
163542..163659
/note="matching EMBL:G20222
RHdb:RH16711
dbSTS:STS825
Identified using the e-PCR software (G. Schuler)"
163594..163754
/note="matching EMBL:R37868
RHdb:RH80977
dbSTS:STS57398
Identified using the e-PCR software (G. Schuler)"
163645..163717
/note="matching EMBL:Z41154
RHdb:RH36666
dbSTS:STS9521
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 49485 a 32505 c 30834 g 52111 t
ORIGIN

Query Match 4.2%: Score 20; DB 9; Length 164935;
Best Local Similarity 100.0%; Pred No. 6.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TCAGAGTGGCAGAGAGGCT 105
|||||
```


Db 119882 TCAGATGGCAGAGAGCT 119863

RESULT 10

AC024429/c

LOCUS

DEFINITION

Homo sapiens chromosome 5 clone RP11-402F5 map 5, WORKING DRAFT

SEQUENCE

AC024429

VERSION

AC024429.2

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1071: contig of 1071 bp in length
 1072 1171: gap of 100 bp
 1172 1257: contig of 86 bp in length
 1258 1357: gap of 100 bp
 1358 2420: contig of 1063 bp in length
 2421 2520: gap of 100 bp
 2521 3552: contig of 1032 bp in length
 3553 3652: gap of 100 bp
 3653 4848: contig of 1196 bp in length
 4849 4948: gap of 100 bp
 4949 6121: contig of 1173 bp in length
 6122 6221: gap of 100 bp
 6222 8087: contig of 1866 bp in length
 8088 8187: gap of 100 bp
 8188 11231: contig of 3044 bp in length
 11232 11331: gap of 100 bp
 11332 15098: contig of 3767 bp in length
 15099 15198: gap of 100 bp
 15199 17028: contig of 1830 bp in length
 17029 17128: gap of 100 bp
 17129 20262: contig of 3134 bp in length
 20263 20362: gap of 100 bp
 20363 23631: contig of 3269 bp in length
 23632 23731: gap of 100 bp
 23732 27168: contig of 3437 bp in length
 27169 27688: gap of 100 bp
 27689 30351: contig of 3083 bp in length
 30352 30451: gap of 100 bp
 30452 33936: contig of 3485 bp in length
 33937 34036: gap of 100 bp
 34037 37015: contig of 2979 bp in length
 37016 37115: gap of 100 bp
 37116 42151: contig of 5036 bp in length
 42152 42251: gap of 100 bp
 42252 47592: contig of 5341 bp in length
 47593 47692: gap of 100 bp
 47693 53063: contig of 5371 bp in length
 53064 53163: gap of 100 bp
 53164 58586: contig of 5423 bp in length
 58587 58686: gap of 100 bp
 58687 64250: contig of 5564 bp in length
 64251 64350: gap of 100 bp
 64351 71393: contig of 7043 bp in length
 71394 71493: gap of 100 bp
 71494 77504: contig of 6011 bp in length
 77505 77604: gap of 100 bp
 77605 85490: contig of 7886 bp in length
 85491 85590: gap of 100 bp
 85591 95960: contig of 10370 bp in length
 95961 96060: gap of 100 bp
 96061 104078: contig of 8018 bp in length
 104079 104178: gap of 100 bp
 104179 114515: contig of 10337 bp in length
 114516 114615: gap of 100 bp
 114616 124221: contig of 9606 bp in length
 124222 124321: gap of 100 bp
 124322 133757: contig of 9436 bp in length
 133758 133857: gap of 100 bp
 133858 148979: contig of 15122 bp in length
 148980 149079: gap of 100 bp
 149080 175291: contig of 26212 bp in length.

FEATURES

Source

1. .175291
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5"
 /clone="RP11-402F5"
 /clone_lib="RPC1-11 Human Male BAC"
 misc_feature 1. .1071

Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6936
 Center clone name: 402.F.5
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 150729 bases at least Q40
 Consensus quality: 163124 bases at least Q30
 Consensus quality: 168741 bases at least Q20
 Insert size: 186000; agarose-fp
 Insert size: 172291; sum-of-contigs
 Quality coverage: 3.3 in Q20 bases; agarose-fp
 Quality coverage: 3.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 31 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

```
misc_feature /note="assembly_fragment"
1172. 1257
/note="assembly_fragment
vector_end:T7
vector_side:right"
1358. 2420
misc_feature /note="assembly_fragment"
2521. 3552
misc_feature /note="assembly_fragment"
3653. 4848
misc_feature /note="assembly_fragment"
4949. 6121
misc_feature /note="assembly_fragment"
6222. 8087
misc_feature /note="assembly_fragment"
8188. 11231
misc_feature /note="assembly_fragment"
11332. 15098
misc_feature /note="assembly_fragment"
15199. 17028
misc_feature /note="assembly_fragment"
17129. 20262
misc_feature /note="assembly_fragment"
20363. 23631
misc_feature /note="assembly_fragment"
23732. 27168
misc_feature /note="assembly_fragment"
27269. 30351
misc_feature /note="assembly_fragment"
30452. 33936
misc_feature /note="assembly_fragment
clone_end:SP6
vector_side:right"
34037. 37015
misc_feature /note="assembly_fragment"
37116. 42151
misc_feature /note="assembly_fragment"
42252. 47592
misc_feature /note="assembly_fragment"
47693. 53063
misc_feature /note="assembly_fragment"
53164. 58586
misc_feature /note="assembly_fragment"
58687. 64250
misc_feature /note="assembly_fragment"
64351. 71393
misc_feature /note="assembly_fragment"
71494. 77504
misc_feature /note="assembly_fragment"
77605. 85490
misc_feature /note="assembly_fragment"
85591. 95960
misc_feature /note="assembly_fragment"
96061. 104078
misc_feature /note="assembly_fragment"
104179. 114515
misc_feature /note="assembly_fragment"
114616. 124221
misc_feature /note="assembly_fragment"

Query Match 4.2%; Score 20; DB 2; Length 175291;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 AATACTATATAGACCATTTT 52
|||||
Db 87053 AATACTATATAGACCATTTT 87034

RESULT 11
AC093290/c 187479 bp DNA 16-AUG-2001
LOCUS HTG
DEFINITION Homo sapiens chromosome 5 clone RP11-47L17, WORKING DRAFT SEQUENCE,
```

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

13 unordered pieces.
AC093290
AC093290.1 GI:15193424
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187479)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 187479)
DOE Joint Genome Institute.
Direct Submission
Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 424594
Center clone name: RPCI-11_47L17

Summary Statistics
Consensus quality: 172550 bases at least Q40
Consensus quality: 181778 bases at least Q30
Consensus quality: 183184 bases at least Q20
Estimated insert size: 192990; agarose-fp estimation
Estimated insert size: 186279; sum-of-contigs estimation
Quality coverage: 10.6 in Q20 bases; agarose-fp estimation
Quality coverage: 10.98 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1052: contig of 1052 bp in length
* 1053 1152: gap of unknown length
* 1153 3200: contig of 2048 bp in length
* 3201 3300: gap of unknown length
* 3301 6058: contig of 2758 bp in length
* 6059 6158: gap of unknown length
* 6159 9304: contig of 3346 bp in length
* 9305 16802: gap of unknown length
* 16803 16902: contig of 7198 bp in length
* 16903 22096: contig of 5194 bp in length
* 22097 22196: gap of unknown length
* 22197 32464: contig of 10268 bp in length
* 32465 32565: gap of unknown length
* 32566 44124: contig of 11560 bp in length
* 44125 44225: gap of unknown length
* 44226 60616: contig of 16392 bp in length
* 60617 60717: gap of unknown length
* 60718 86562: contig of 25846 bp in length
* 86563 86663: gap of unknown length
* 86664 111269: contig of 24607 bp in length
* 111270 144754: gap of unknown length
* 144755 144854: contig of 33385 bp in length
* 144855 187479: contig of 42625 bp in length.
Location/Qualifiers
1. 187479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-47L17"

FEATURES
SOURCE

BASE COUNT 57600 a 35701 c 35441 g 57486 t 1251 others
 ORIGIN

Query Match 4.2%; Score 20; DB 2; Length 187479;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 AATACCTATAGAGCCATTTT 52

|||||

Db 1581 AATACCTATAGAGCCATTTT 1562

RESULT 12

AC022739

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 190174)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 190174)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,

Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Penestor,J.,

Ferrellano,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,

Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,

MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,

Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,

Zimmer,A. and Zody,M.

Direct Submission

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Oct 4, 2000 this sequence version replaced gi:7960374.

All repeats were identified using RepeatMasker.

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5613

Center clone name: 570.L13

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 185540 bases at least Q40

Consensus quality: 187787 bases at least Q30

Consensus quality: 188645 bases at least Q20

Insert size: 186000; agarose-fp

Insert size: 189574; sum-of-contigs

Quality coverage: 7.1 in Q20 bases; agarose-fp

Quality coverage: 7.0 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 58818: contig of 58818 bp in length
 * 58819 58918: gap of 100 bp
 * 58919 59926: contig of 1008 bp in length
 * 59927 60026: gap of 100 bp
 * 60027 62638: contig of 2612 bp in length
 * 62639 62738: gap of 100 bp
 * 62739 73191: contig of 10453 bp in length
 * 73192 73291: gap of 100 bp
 * 73292 90983: contig of 17692 bp in length
 * 90984 91083: gap of 100 bp
 * 91084 145445: contig of 54362 bp in length
 * 145446 145545: gap of 100 bp
 * 145546 190174: contig of 44629 bp in length.

FEATURES

Location/Qualifiers

1..190174

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-570L13"

/clone_lib="RPC1-11 Human Male BAC"

1..58818

/note="assembly_fragment"

vector_side:left"

58919..59926

/note="assembly_fragment"

60027..62638

/note="assembly_fragment"

62739..73191

/note="assembly_fragment"

73292..90983

/note="assembly_fragment"

91084..145445

/note="assembly_fragment"

145546..190174

/note="assembly_fragment"

vector_side:right"

59581 a 37142 c 36772 g 56069 t

BASE COUNT

ORIGIN

Query Match 4.2%; Score 20; DB 2; Length 190174;

Best Local Similarity 100.0%; Pred. No. 6.7;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 TTCTTAAGGGGAAATTAATC 126

|||||

Db 26355 TTCTTAAGGGGAAATTAATC 26374

RESULT 13

AC087841

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC087841

192725 bp DNA HTG

Mus musculus chromosome 6 clone RP23-415I19 strain C57BL6/J,

WORKING DRAFT SEQUENCE, 19 unordered pieces.

AC087841

GI:12621392

HTG; HTGS_PHASE1; HTGS_DRAFT.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192725)

Ayelle, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,
Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,
Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B., Masiello, C.,
Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
Shevchenko, Y., Snyder, B., Stantripop, S., Thomas, J.W., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Wetherby, K.D., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished

2 (bases 1 to 192725)
Green, E.D.
Direct Submission

Submitted (31-JAN-2001) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA

----- Genome Center
Center: NIH Intramural Sequencing Center

Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.mouse@hri.nih.gov

----- Project Information
Center project name: vg
Center clone name: 415119

----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319
Consensus quality: 181855 bases at least Q40
Consensus quality: 184839 bases at least Q30
Consensus quality: 186657 bases at least Q20

Insert size: 192000; agarose-fp
Insert size: 217000; pulse-field-gel
Insert size: 190925; sum-of-contigs

Quality coverage: 5.96x in Q20 bases; agarose-fp
Quality coverage: 5.27x in Q20 bases; pulse-field-gel
Quality coverage: 5.99x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 19 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 2023: contig of 2023 bp in length
2024 2123: gap of unknown length
2124 4491: contig of 2368 bp in length
4492 4591: gap of unknown length
4592 7439: contig of 2848 bp in length
7440 7539: gap of unknown length
7540 10623: contig of 3084 bp in length
10624 10723: gap of unknown length
10724 14491: contig of 3768 bp in length
14492 14591: gap of unknown length
14592 18465: contig of 3873 bp in length
18466 18565: gap of unknown length
18566 23231: contig of 4667 bp in length
23232 23332: gap of unknown length
23333 30382: contig of 7051 bp in length
30383 30482: gap of unknown length
30483 38702: contig of 8220 bp in length
38703 38802: gap of unknown length
38803 46799: contig of 7997 bp in length
46800 46899: gap of unknown length
46900 56572: contig of 9673 bp in length
56573 56672: gap of unknown length
56673 65347: contig of 8675 bp in length
65348 65448: gap of unknown length
65449 75185: contig of 9738 bp in length
75186 75285: gap of unknown length

* 75286 86139: contig of 10854 bp in length
* 86140 86239: gap of unknown length
* 86240 99372: contig of 13133 bp in length
* 99373 99472: gap of unknown length
* 99473 117995: contig of 18523 bp in length
* 117996 118095: gap of unknown length
* 118096 136106: contig of 18011 bp in length
* 136107 136206: gap of unknown length
* 136207 162352: contig of 26146 bp in length
* 162353 162452: gap of unknown length
* 162453 192725: contig of 30273 bp in length.

FEATURES
source
1..192725
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="6"
/clone_lib="RPCI mouse BAC library 23"
1..2023
misc_feature
/note="assembly_fragment"
2124..4491
misc_feature
/note="assembly_fragment"
4592..7439
misc_feature
/note="assembly_fragment"
7540..10623
misc_feature
/note="assembly_fragment"
10724..14491
misc_feature
/note="assembly_fragment"
14592..18464
misc_feature
/note="assembly_fragment"
18565..23231
misc_feature
/note="assembly_fragment"
23332..30382
misc_feature
/note="assembly_fragment"
30483..38702
misc_feature
/note="assembly_fragment"
38803..46799
misc_feature
/note="assembly_fragment"
46900..56572
misc_feature
/note="assembly_fragment"
56673..65347
misc_feature
/note="assembly_fragment"
clone_end:r7
vector_side:left
65448..75185
misc_feature
/note="assembly_fragment"
75286..86139
misc_feature
/note="assembly_fragment"
86240..99372
misc_feature
/note="assembly_fragment"
99473..117995
misc_feature
/note="assembly_fragment"
118096..136106
misc_feature
/note="assembly_fragment"
136207..162352
misc_feature
/note="assembly_fragment"
162453..192725
misc_feature
/note="assembly_fragment"

BASE COUNT 52082 a 43032 c 43919 g 51808 t 1884 others
ORIGIN

Query Match 4.2% Score 20; DB 2: Length 192725;
Best Local Similarity 100.0%; Pred.No. 6.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 449 TTTCAGAGAGAGAACTC 468
|||||
DB 11299 TTTCAGAGAGAGAACTC 11318

RESULT 14

HIVU69075
LOCUS
DEFINITION HIV-1 patient D clone L5 from USA reverse transcriptase (pol) gene, partial cds.
ACCESSION U69075.1 GI:1546647
VERSION
KEYWORDS
SOURCE
ORGANISM Human immunodeficiency virus type 1.
Human immunodeficiency virus type 1.
Viruses; Retroviral viruses; Retroviridae; Lentivirus; Primate lentivirus group.
REFERENCE 1 (bases 1 to 380)
AUTHORS Wong, J.K.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1996) Department of Medicine, University of California School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093-0679, USA

FEATURES
source Location/Qualifiers
1..380
/organism="Human immunodeficiency virus type 1"
/isolate="patient D"
/db_xref="taxon:11676"
/tissue_type="lymph node"
/clone="L5"
/note="from San Diego, USA"
1..380
/gene="pol"
1..380
/gene="pol"
/note="encodes terminal 8 codons of the protease followed by codons 1-118 of the reverse transcriptase as represented in HIV-1 LAI GenBank Accession Number K02013"
/codon_start=1
/product="reverse transcriptase"
/protein_id="BAB08364.1"
/db_xref="GI:1546648"
/translation="QIGCTLNFPIPIETVPVKLPKGMGPKVKOMPLTEEKIKALIE ICTMEKEGKGVSKVPENPTVPFAIRKDKSTWRKIVDFRELNKKTDQFEVQLGHPHAGLKKKSVTVLDVGDVAFVSP"

BASE COUNT 157 a 54 c 74 g 95 t
ORIGIN

Query Match 4.0%; Score 19; DB 14; Length 380;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 TGGAAAGGAGGAAAGT 149
|||||
DB 146 TGGAAAGGAGGAAAGT 164

RESULT 15
HSP48B07
LOCUS
DEFINITION H.sapiens flow-sorted chromosome 1 HindIII fragment, SC1p48B07, sequence tagged site.
ACCESSION AL033810
VERSION AL033810.1 GI:3893549
KEYWORDS STS; Anonymous marker; single read.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS Gregory, S., Kettleborough, R., Langford, C., Ross, M.T. and Hunt, S.E.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1998) The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact: humquery@sanger.ac.uk
Vector: pBSiISK+
Marker stSG3217FS (Primer A : AGAAGAAATGAAGGCATCA; Primer B : CCTGCAACTTCACTGAATTCa; amplicon size : 139 bp) was mapped to

chromosome 1 using Radiation Hybrid panel Genebridge 4 (GB4).
Location/Qualifiers
1..418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/sex="Female"
/dev_stage="adult"
/tissue_type="EBV lymphoblastoid cell line"
/clone_lib="SC1p48B07"
/clone="SC1p48B07"

BASE COUNT 156 a 78 c 82 g 102 t
ORIGIN

Query Match 4.0%; Score 19; DB 11; Length 418;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 GAGCAGTTTAGCAAGAGAA 255
|||||
DB 91 GAGCAGTTTAGCAAGAGAA 109

Search completed: January 31, 2002, 12:14:33
Job time: 1506 sec

OM of: US-08-957-709-70 to: Issued_Patents_AA:* out_format : pfs
 Date: Jan 31, 2002 12:58 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=framet.n2p.model -DEV=xlh
 -Q/cgn2_1/USPTO.spool/US08957709/runat_30012002.133258.9490/app_query.fasta_1.536
 -DB-Issued_Patents_AA -OFMT=fastan -SUFFIX=rai -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US08957709.@cgn1_1.0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-957-709-70
 Query length: 471
 Database: Issued_Patents_AA:*
 Database sequences: 212252
 Database length: 22503292
 Search time (sec): 26.380000

score_list:	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-45 +	558.00	1279.26	3.4e-64	246	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-44 +	533.00	1220.78	6.1e-61	246	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-43 +	436.00	993.90	2.7e-48	246	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-41 +	160.00	369.20	9.8e-13	42	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-37 +	124.00	290.21	3.8e-08	28	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-39 +	124.00	289.78	3.9e-08	27	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-38 +	119.00	278.51	1.7e-07	27	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-11 +	114.50	248.74	1.5e-06	138	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-2 +	112.50	242.03	3.0e-06	164	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-4 +	112.50	236.96	3.8e-06	252	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-10 +	112.00	242.90	3.2e-06	138	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-238-303-9 +	87.50	160.58	0.0149	1150	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-12 +	87.00	183.68	0.0060	147	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-13 +	84.50	171.05	0.0310	144	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-6 +	79.50	166.13	0.0571	147	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-307-499-22 +	78.50	164.20	0.0757	142	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-299-268-22 +	78.50	164.20	0.0757	142	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-54 +	76.50	186.85	0.0420	14	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-258-639A-4 +	76.50	141.42	0.3029	659	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-900-951-4 +	76.50	141.42	0.3029	659	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-11 +	74.00	178.71	0.0983	17	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-7 +	73.50	151.78	0.3500	151	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-453-472-8 +	72.00	136.10	0.9320	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-038-948-2 +	72.00	136.10	0.9320	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-038-948-5 +	72.00	136.10	0.9320	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-453-952-8 +	72.00	136.10	0.9320	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-862-903-8 +	72.00	136.10	0.9320	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-8 +	71.00	146.74	0.7153	141	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-824-405-9 +	71.00	146.66	0.7179	142	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-453-472-7 +	71.00	133.76	1.26	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-038-948-4 +	71.00	133.76	1.26	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-453-952-7 +	71.00	133.76	1.26	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-862-903-7 +	71.00	133.76	1.26	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-484-158B-61 +	71.00	133.76	1.26	424	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-484-993B-8 +	70.50	132.84	1.45	415	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-484-158B-8 +	70.50	132.84	1.45	415	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-484-596A-8 +	70.50	132.84	1.45	415	
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-480-150A-8 +	70.50	132.84	1.45	415	
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-458-731-8 +	70.50	132.84	1.45	415	

/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-149-223A-8 + 70.50 132.84 1.45 415
 /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-374-483-2 + 70.00 123.64 2.38 820
 /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-374-483-5 + 70.00 123.64 2.38 820
 /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-599-171A-25 + 69.00 129.36 2.26 414
 /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-646-590B-25 + 69.00 129.36 2.26 414

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-45

seq_documentation_block:
 : Sequence 45, Application US/08822774
 : Patent No. 6183997
 : GENERAL INFORMATION:
 : APPLICANT: HOGREFE, Holly
 : TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
 : TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
 : TITLE OF INVENTION: and Methods for Purifying and Identifying Same
 : NUMBER OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
 : ADDRESS: Lenahan, P.L.L.C.
 : STREET: 1200 G Street, N.W. Suite 700
 : CITY: Washington
 : STATE: D.C.
 : ZIP: 20005

: COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/822,774
 : FILING DATE: 21-MAR-1997
 : CLASSIFICATION: 436

: ATTORNEY/AGENT INFORMATION:
 : NAME: KULIK, David J.
 : REGISTRATION NUMBER: 36,576
 : REFERENCE/DOCKET NUMBER: 1486/43163

: TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 628-8800
 : TELEFAX: (202) 628-8844
 : INFORMATION FOR SEQ ID NO: 45:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 246 amino acids
 : TYPE: amino acid
 : TOPOLOGY: unknown
 : MOLECULE TYPE: peptide
 : HYPOTHEetical: YES

US-08-822-774-45

alignment_scores:
 Quality: 558.00 Length: 105
 Ratio: 5.417 Gaps: 0
 Percent Similarity: 98.095 Percent Identity: 98.095

alignment_block:
 US-08-957-709-70 x US-08-822-774-45

Align seg 1/1 to: US-08-822-774-45 from: 1 to: 246

156 TCCTCAAGGAATACGCTTAATCCTAACCCCTCGAGAGATAAGTTGC 205
 |||||
 1 SerSerLysGlyIleArgLeuAsnProAsnProArgLeuAspLysValal 17
 206 CGACCATGTTATGGGGATATGAAGATAAGCAGTGTAGCAAGAA 255
 |||||
 17 aArgArgCysTyrGlyGlyTyrGluAspLysGluGlnPheSerLysArg 34
 256 GGGGTATTGTTCTTTTCTTGGTGTGCTGACCCAGCATGGATGAACTT 305
 |||||
 34 rgGlyTyrTrpPhePheCysLeuGly***ProArgMetGlyTrpLysLeu 50
 306 AACACTAATGCTCTACAATGCTCAATGAACCTCGAATTAAGATATG 355

```

|||||
51 AsnThrAsnAlaLeuGlnCysLeuLys***ThrCysArgIleLysIletr 67
|||||
356 GAGAGAGATTGTGCAGATCGCATTTATTAAGCTAGAGGTCCGGCAAGA 405
|||||
67 PARGLIuileCysAlaAspArgIleIryLysAlaArgGlySerGlyLysL 84
|||||
406 ACCCTTACAGAGGAACATATCAGGGGACACAAAGTTAGCGTTTTCAAA 455
|||||
84 yProLeuGlnArgLysLeuSerGlyGluHisLysValSerValPheLys 100
|||||
456 GAGAAAGAAACTCTA 470
|||||
101 GluLysGluThrLeu 105

```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-822-774-44

seq_documentation_block:
; Sequence 44, Application US/08822774

```

; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/822,774
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHEICAL: YES
US-08-822-774-44

```

alignment_scores:
Quality: 533.00 Length: 104
Ratio: 5.125 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.077

alignment_block:

US-08-957-709-70 x US-08-822-774-44 ..

Align seg 1/1 to: US-08-822-774-44 from: 1 to: 246

```

157 CCTCCAGGGAATACGCTTAATCCCTACCTCGAGAGGATAAAGTGGCC 206
|||||
1 ProProArgGluTyrAlaLeuIleuThrLeuGluArgIleLysLeuPr 17

```

```

207 CGACGATGTTATGGGGATATCAAGATTAAGGAGCAGTTTAGCAAGAGAG 256
|||||
17 OasnAsnValMetGlyAspMetLysIleArgSerLeuAlaArgGluG 34
|||||
257 GGGTTATTGCTTCTTCTGCTGGTTGACCCAGAGATGGATGGAACCTTA 306
|||||
34 lyValIleGlySerPheAlaIrrpValAspProGlyTrrpAspGlyAsnLeu 50
|||||
307 ACACATAATGCTCTACAATGCCTCAATGAACCTGTCGAATTAAAGATATGG 356
|||||
51 ThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeuArgTyrGl 67
|||||
357 AGAGAGATTTGTGCAGATCGCATTTTATAAGGCTAGAGGTCGCGCAAGAA 406
|||||
67 yGluArgPheValGlnIleAlaPheIleArgLeuGluGlyProAlaArg 84
|||||
407 ACCCTTACAGAGGAACATATCAGGGGACACAAAGTTAGCGTTTTCAAA 456
|||||
84 snProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPheSerLys 100
|||||
457 AGAAAGAAACTC 468
|||||
101 ArgLysLysLeu 104

```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-822-774-43

seq_documentation_block:

; Sequence 43, Application US/08822774

; Patent No. 6183997

; GENERAL INFORMATION:

; APPLICANT: HOGREFE, Holly

; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)

; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,

; TITLE OF INVENTION: and Methods for Purifying and Identifying Same

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &

; ADDRESSEE: Lenahan, P.L.L.C. Suite 700

; STREET: 1200 G Street, N.W.

; CITY: Washington

; STATE: D.C.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/822,774

; FILING DATE: 21-MAR-1997

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: KULIK, David J.

; REGISTRATION NUMBER: 36,576

; REFERENCE/DOCKET NUMBER: 1486/43163

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 628-8800

; TELEFAX: (202) 628-8844

; INFORMATION FOR SEQ ID NO: 43:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 246 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

; HYPOTHEICAL: YES

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

US-08-822-774-43

alignment_scores:
Quality: 436.00 Length: 104


```
Ratio: 4.739          Gaps: 0
Percent Similarity: 88.462  Percent Identity: 88.462
alignment_block:
US-08-957-709-70 x US-08-822-774-43
Align seg 1/1 to: US-08-822-774-43 from: 1 to: 246
158 CTCCAAGGAATACCCCTAATCTTAACCCGAGGAGGATGAAGTTGCC 207
|||||
1 LeuGlnGlyAsnThrPro***Ser***ProSerArgGly***SerCysApr 17
208 GACGATGTTATGGGGATATCAAGATAGGACAGCTTTAGCAAGAGAAGG 257
|||||
17 ofthrMetLeuTrpGlyIle***Arg***GlyAlaVal**GlnGluLysG 34
258 GGTATTGGTCTTTTGGCTGGGTGACCCAGGATGGATGGAACCTAA 307
|||||
34 IyLeuLeuValLeuLeuLeuGlyLeuThrGlnAspGlyMetGluThr*** 50
308 CACTAATGCTCTACAACTGCTCAATGAACCTGTCGAATTAGATATGGA 357
|||||
51 HIs***CysSerThrMetProGlnMetAsnLeuSerAsn***AspMetG 67
358 GAGAGATTGTGCAGATCGCATTTTATAGGCTAGAGGTCGCGCAAGAAA 407
|||||
67 uArgAspLeuCysArgSerHisLeu***Gly***ArgValArgGlnGluT 84
408 CCCTTACAGAGAACTATCAGGGGAGCACAAAGGTTAGCGTTTCAAGA 457
|||||
84 hrLeuThrGluGluThrIleArgGlyAlaGlnGly***ArgPheGlnArg 100
458 GAAAGAACTCT 469
101 GluArgAsnSer 104
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-41
seq_documentation_block:
; Sequence 41, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
```

```
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-822-774-41
alignment_scores:
Quality: 160.00 Length: 32
Ratio: 5.161 Gaps: 0
Percent Similarity: 96.875 Percent Identity: 96.875
alignment_block:
US-08-957-709-70 x US-08-822-774-41
Align seg 1/1 to: US-08-822-774-41 from: 1 to: 42
1 ATGCTACTCCAGACTGGAATAATCAGAAAGAAATACTTATAGAGCCATT 50
|||||
11 MetLeuLeuProAspTrpIleArgLysGluIleLeuGluProph 27
51 TTCTGAAGAATGCTCCCAACCCAGCAGGTATGACCTCAGAGTGGC 96
|||||
27 eSerGluGluTrpLeuGlnProAlaGlyTyAspLeuArgValGly 42
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-822-774-37
seq_documentation_block:
; Sequence 37, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
US-08-822-774-37
```

```

alignment_scores:
  Quality: 124.00      Length: 27
  Ratio: 5.167        Gaps: 0
  Percent Similarity: 88.889  Percent Identity: 88.889

alignment_block:
US-08-957-709-70 x US-08-822-774-37 ..
Align seg 1/1 to: US-08-822-774-37 from: 1 to: 27

89 GAGTGGGCAGAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAG 138
|||||
1 GlutPalaGluArgLeuLeuArgGlyAsn***SerLysTrpLysar 17
|||||

139 GAAGGAAAAGTCGTTATCTCCCAAGGGAAT 169
|||||
17 glysGluLysSer***PheLeuGlnGlyasn 27
|||||

seq_name: /cgn2_5/ptodata/2/1aa/6B_COMB.pep:US-08-822-774-39

seq_documentation_block:
; Sequence 39, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/822,774
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8844
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-08-822-774-39

```

```

alignment_scores:
  Quality: 124.00      Length: 28
  Ratio: 4.960        Gaps: 0
  Percent Similarity: 89.286  Percent Identity: 85.714

alignment_block:
US-08-957-709-70 x US-08-822-774-39 ..
Align seg 1/1 to: US-08-822-774-39 from: 1 to: 28

```

```

87 CAGAGTGGGCAGAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAG 136
|||||
1 GlnSerGlyLysArgCys***GlyGluLeuAsnArgSerGlyLys 17
|||||

137 AGGAAGAAAAGTCGTTATCTCCCAAGGGAATA 170
|||||
17 sglYArgLysSerArgTyrSerSerLysGlyLeu 28
|||||

seq_name: /cgn2_5/ptodata/2/1aa/6B_COMB.pep:US-08-822-774-38

seq_documentation_block:
; Sequence 38, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/822,774
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8844
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-08-822-774-38

```

```

alignment_scores:
  Quality: 119.00      Length: 27
  Ratio: 4.760        Gaps: 0
  Percent Similarity: 92.593  Percent Identity: 88.889

alignment_block:
US-08-957-709-70 x US-08-822-774-38 ..
Align seg 1/1 to: US-08-822-774-38 from: 1 to: 27

```

```

88 AGAGTGGGCAGAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAG 137
|||||
1 ArgValGlyArgGluAla***ValLysGlyLysLeuIleGluValGluLys 17
|||||

138 GAAGGAAAAGTCGTTATCTCCCAAGGGA 168
|||||
17 sgluglyLysVal***IleProProArgGlu 27
|||||

```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pap:US-08-824-405-11

seq_documentation_block:
; Sequence 11, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladhner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna Salvatore J.
; TITLE OF INVENTION: Outpase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-824-405-11

alignment_scores:
Quality: 114.50 Length: 123
Ratio: 1.568 Gaps: 1
Percent Similarity: 59.350 Percent Identity: 22.764

alignment_block:
US-08-957-709-70 x US-08-824-405-11 ..

Align seg 1/1 to: US-08-824-405-11 from: 1 to: 138

55 GAAGATCGTCCACACAGCAGTGTATACCTCAGCAGTGGCAGAGGC 104
||||| : : : ||||| ||| : : : ||||| : : :
14 GluLysArgGluAspAlaGlyPheAspLeuGlyValProTyrAsp.. 29
||| : : : ||||| : : : ||||| : : : ||||| : : :
105 TTTTCTTAAGGGAAATTAATCGACGTGAAAGGAAAGGAAAGTCGTTA 154
: : : : :
30IleMetI 32
155 TTCTTCAAGGAATACGCTTAATCTTAACCTCGAGAGATAAGTTG 204
||||| : : : ||||| ||| : : : ||||| : : :
32 leProValSerAspThrLysIlelleProThrAspValIleVal 48
205 CCCGACGATGTTATGGGGATATGAAGATAAGACAGTTCAGCAAGAGA 254
||| : : : ||||| : : : ||||| : : : ||||| : : :
49 ProProAsnSerPheGlyTyrValThrGlyLysSerSerMetAlaLysGI 65

255 AGGGTTATTGTTCTTTTGGTTGACCCAGGATGGGATGAAACT 304
: ||| : : : : : ||||| ||||| : : : ||||| : : :
65 nGlyLeuLeuIleAsnGlyGlyIleIleAspGluGlyTyrThrGlyGluI 82
: : : : : ||| : : : : : : : : : : : : : : :
305 TAACACTAATGCTCTACATGCTCAAAATGAACCTGCGAATTAAGATAT 354
: : : : : ||| : : : : : : : : : : : : : : :
82 leGlnValIleCysThrAsnIleGlyLysSerAsnIleLysLeuIleGlu 98
: : : : : ||| : : : : : : : : : : : : : : :
355 GGAGAGAGATTTGTCAGATGCCATTATAGCGTAGAGGGTCCGGCAAG 404
||| : : : ||||| : : : : : : : : : : : : : : :
99 GlyGlnLysPheAlaGlnLeuIleIleLeuGlnHisHisSerAsnSerAr 115
: : : : : ||| : : : : : : : : : : : : : : :
405 AAACCCCTTACAGAGGAAAC 423
| : : : ||||| : : : |||
115 gGlnProTyrAspGluAsn 121
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pap:US-08-824-405-2
seq_documentation_block:
; Sequence 2, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladhner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: Outpase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-824-405-2

alignment_scores:
Quality: 112.50 Length: 132
Ratio: 1.607 Gaps: 4
Percent Similarity: 53.030 Percent Identity: 29.545

alignment_block:
US-08-957-709-70 x US-08-824-405-2 ..

```
Align seg 1/1 to: US-08-824-405-2 from: 1 to: 164
61 TCGCTCCACACGAGGTATGACCTCAGCTGAGTGGCAGAGAGCGTTTGT 110
   |||  ::  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
42 SerAlaArgAlaGlyTyrAspLeuTyrSerAlaTyrAsp..... 55
   |||  ::  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 TAAGGGAAATTAATCGACGTGGAAGGAAAGGAAAGTCGTTATTCCTC 160
   |||  ::  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
56 .....TyrThrIleProp 60
161 CAAGGAATACCCCTTAATCCTAACCTCAGAGGATTAAGTTCGCCGAC 210
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
60 roMetGluLysAlaValValIleThrAspIleGlnIleAlaLeuProSer 76
211 GATCTATGGGGATATAGATAGGACGAGTTCAGCAAGAGAGGGGT 260
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
77 GlyCysTyrGlyArgValAlaProArgSerGlyLeuAlaAlaLysHisph 93
261 TATTCGTTCTTTTCTGGT...GACCCAGGATGGGATGGAACCTTAA 307
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
93 eileAspValGlyAlaGlyValIleAspGluAspTyrArgGlyAsnValG 110
308 CACTAATGCTCACAATGCTCAATGACCTGTCGAATTAAGATATGGA 357
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
110 lyValValLeuPheAsnPheGlyLysGluLysPheGluValLysGly 126
358 GAGAGATTGTCAGATCGCATTTATAAGGCTA.....GAGGGTCC 398
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
127 AspArgIleAlaGlnLeuIleCysGluArgIlePheTyrProGlu.IleG 143
399 GCACAGAAACCCCTACAGAGGAACTATCAGGGGAGCAGCAAGGT 442
143 luGluValGlnAlaLeuAspThrGluArgGlySerGlyGly 157
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-824-405-4
seq_documentation_block:
; Sequence 4, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: OutPhase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
```

```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-824-405-4
alignment_scores:
  Quality: 112.50      Length: 132
  Ratio: 1.607        Gaps: 4
  Percent Similarity: 53.030  Percent Identity: 29.545
alignment_block:
US-08-957-709-70 x US-08-824-405-4 ..
Align seg 1/1 to: US-08-824-405-4 from: 1 to: 252
61 TCGCTCCACACGAGGTATGACCTCAGCTGAGTGGCAGAGAGCGTTTGT 110
   |||  ::  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 SerAlaArgAlaAlaGlyTyrAspLeuTyrSerAlaTyrAsp..... 143
   |||  ::  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 TAAGGGAAATTAATCGACGTGGAAGGAAAGGAAAGTCGTTATTCCTC 160
   |||  ::  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 .....TyrThrIleProp 148
161 CAAGGAATACCCCTTAATCCTTAACCTCAGAGGATTAAGTTCGCCGAC 210
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
148 roMetGluLysAlaValValIleThrAspIleGlnIleAlaLeuProSer 164
211 GATCTATGGGGATATAGATAGGACGAGTTCAGCAAGAGAGGGGT 260
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
165 GlyCysTyrGlyArgValAlaProArgSerGlyLeuAlaAlaLysHisph 181
261 TATTCGTTCTTTTCTGGT...GACCCAGGATGGGATGGAACCTTAA 307
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
181 eileAspValGlyAlaGlyValIleAspGluAspTyrArgGlyAsnValG 198
308 CACTAATGCTCACAATGCTCAATGACCTGTCGAATTAAGATATGGA 357
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
198 lyValValLeuPheAsnPheGlyLysGluLysPheGluValLysGly 214
358 GAGAGATTGTCAGATCGCATTTATAAGGCTA.....GAGGGTCC 398
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
215 AspArgIleAlaGlnLeuIleCysGluArgIlePheTyrProGlu.IleG 231
399 GCACAGAAACCCCTACAGAGGAACTATCAGGGGAGCAGCAAGGT 442
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
231 luGluValGlnAlaLeuAspThrGluArgGlySerGlyGly 245
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-824-405-10
seq_documentation_block:
; Sequence 10, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: OutPhase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 138 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-824-405-10

alignment_scores:
  Quality: 112.00      Length: 133
  Ratio: 1.534        Gaps: 3
  Percent Similarity: 54.887      Percent Identity: 26.316

alignment_block:
US-08-957-709-70 x US-08-824-405-10 ..
  Align seg 1/1 to: US-08-824-405-10 from: 1 to: 138
55 GAAGAATCGCTCCCAACGAGAGGTTATGACCTCAGAGTGGGAGAGGCG 104
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
14 GlnLysArgAlaGluAspAlaGlyTyrAspLeuIleCysProGlnGlu.. 29
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
105 TTTTGTAAAGGGAATTAATGAGCTGGAAAGGAGAAAGTCGTTA 154
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
30 .....IleSerI 32

155 TTCTCCCAAGGAATACGCTTAATCTTAACCTCAGAGGAGTAAGTTG 204
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
32 IeProlaGlyGlnValLysArgIleAlaIleAspLeuLysIleAsnLy 48
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
205 CCGGACGATGTTATGGGGATGATGAAGATAAGGAGCAGTTAGCAAGAGA 254
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
49 LysLysAspGlnTrrAlaMetIleGlyThrLysSerSerPheAlaAsnLy 65
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
255 AGGGTTATTGTTCTTTTGTGTTGGTTGACCCAGGATGGGATGAAACT 304
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
65 scLyValPheValGlnGlyIleIleAspSerGlyTyrGlnGlyThrI 82
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
305 TAACTAATGCTCTACAAATGACCTCAATGACCTGCTGAATTAAGATAT 354
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
82 IeGlnValIleTyrAsnSerAsnAsnLysGluValIleProGln 98
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
355 GGAGAGAGATTGTGCAGATGCATTTTATAAGGCTAGAGGGTCCGGCAAG 404
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
99 GlyArgLysPheAlaGlnIleLeuMetProLeuIleHisGluGluLe 115
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
405 AAACCTTAC.....AGAGAAACTATCAGGGGAGCACAAGGT 442
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
115 uCluProTrpGlyGluThrArg.LysThrGluArgGlyGluGlnGly 130
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.US-09-238-303-9
seq_documentation_block:
; Sequence 9, Application US/09238303B
; Patent No. 6284253

; GENERAL INFORMATION:
; APPLICANT: Barr, Margaret C.
; TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequenc
; FILE REFERENCE: 18617.0059
; CURRENT APPLICATION NUMBER: US/09/238,303B
; CURRENT FILING DATE: 1999-01-28
; EARLIER APPLICATION NUMBER: US 60/072,927
; EARLIER FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 9
; LENGTH: 1150
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: protein encoded by the pol gene of a recombinant viral
; OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat felin
; OTHER INFORMATION: immunodeficiency virus
US-09-238-303-9

alignment_scores:
  Quality: 87.50      Length: 108
  Ratio: 1.535        Gaps: 3
  Percent Similarity: 52.778      Percent Identity: 24.074

alignment_block:
US-08-957-709-70 x US-09-238-303-9 ..
  Align seg 1/1 to: US-09-238-303-9 from: 1 to: 1150
67 CAACGACGAGGTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGG 116
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
750 GluAspAlaGlyTyrAspLeuAla.....758

117 GAAATTAATCGACGTGGAAAGGAAAGAAAGTCTTATTCTCCAAGGG 166
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
759 .....GluGlnAspIleHisLeuMetProGlyG 768

167 AATAGCCCTTAATCCTAACCTCAGAGAGATAAAGTTGCCCGCAGCATGT 216
   ||:::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
768 luValArgIleValProThrGlyValArgLeuMetLeuProLysGlyHis 784
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
217 ATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGAAGG.....GT 260
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
785 TrpGlyMetValValGlyLysSerSerIleAlaLysGlnGlyLeuAspVa 801
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
261 TATTGGTTCTTTTGTGTTGGTTGACCCAGGATGGGATGAAACTTAACAC 310
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
801 lleuGlyGlyVal.....IleAspGluGlyTyrArgGlyGluIleGlyV 816
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
311 TAATGCTCTACAACTCAATGAACCTGTCGAATTAAGATATGGAGAG 360
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
816 alileMetIleAsnLeuGlnLysArgSerIleThrLeuLysGluLysGln 832
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
361 AGATTGTGCAGATGCATTTATA 384
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
833 LysValAlaGlnLeuIleIle 840

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.US-08-824-405-12
seq_documentation_block:
; Sequence 12, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: catFase, its isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
```

```

; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 147 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-824-405-12

alignment_scores:
  Quality: 87.00      Length: 84
  Ratio: 1.673       Gaps: 0
  Percent Similarity: 61.905   Percent Identity: 27.381

alignment_block:
US-08-957-709-70 x US-08-824-405-12  ..

Align seg 1/1 to: US-08-824-405-12 from: 1 to: 147

151 GTATTCTCCCAAGGGAATACGCTTAATCTTAACCTCGAGAGGATAAA 200
|||||  ::  :::::|||||
40 ValIleProSerHisCysSerArgValValPheThrAspLeuLeuIlely 56

201 GTTCCCGCAGCATGTTATGGGGGATAGAGATAGAGCAGCTTTAGCAA 250
| |||::: ||| ::  :::::|||||
56 sProProSerGlyCysTyrGlyArgIleAlaProArgSerGlyAlaValL 73

251 GAGAGGGGTTATTGTTCTTTGTTGGTTGACCCAGGATGGATGGA 300
::: ::  ::: ::  ::: |||
73 yHisPheIleAspValGlyAlaGlyValIleAspGluAspTyrArgGly 89

301 AACTTAACACTAATCTCTACAAATGCCCTCAATGAACCTGTGCAATTAAG 350
|||::: :::::|||||:|||||
90 AsnValGlyValIleValPheAsnPheGlyAsnSerAspPheGluVally 106

351 ATATGAGAGAGATTGTGCAGATCGCATTTATAGGCTAGAGGTCGG 400
::: |||:::|||||:|||||
106 slyGlyAspArgIleAlaGlnLeuIleCysGluArgIleSerCysProA 123

401 CA 402
123 la 123

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.us-08-824-405-13

seq_documentation_block:
; Sequence 13, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:

```

```

; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-824-405-13

alignment_scores:
  Quality: 81.50      Length: 116
  Ratio: 1.358       Gaps: 1
  Percent Similarity: 51.724   Percent Identity: 24.138

alignment_block:
US-08-957-709-70 x US-08-824-405-13  ..

Align seg 1/1 to: US-08-824-405-13 from: 1 to: 144

52 TCTGAAGATCGCTCCAAACCGAGGTTATGACCTCAGATGGCGAGAGA 101
::: ::: ||| ||| ||| ||| ||| ||| ||| :::
20 ThrArgGlnSerProTyrAlaAlaGlyTyrAspLeuTyrSerAlaIlyrAs 36

102 GCCTTTTCTTAGGGGAAATTANTCGACGTGGAAAGGAAAGTCG 151
:
36 P.....TyrT 38

152 TTATTCCTCCAAAGGAATACGCTTAATCTTAACCTCGAGAGGATAAG 201
::: ||| ||| ||| ||| ||| ||| :::
38 hrIlePheProGlyGluArgGlnLeuIleLysThrAspIleSerMetSer 54

202 TTGCCCGCAGCATGTTATGGGGGATATGAAGATAGGACGAGTTTACCAAG 251
::: ||| ||| ::  ::: ||| ||| ||| |||
55 MetProLysGlyCysTyrGlyArgIleAlaProArgSerGlyLeuSerLe 71

252 AGAAGGGGTTATTGTTCTTTTGGTTGGTTGACCCAGGATGGGATGAA 301
::: ||| ::  ::: ||| ||| ||| |||
71 ulysGlyIleAspIleGlyGlyValIleAspGluAspTyrArgGlyA 88

302 ACTTAACACTAATGCTCTACAATGCCTCAATGACCTGTCGAATTAAGA 351

```


OM of: US-08-957-709-70 to: A_Geneseq_1101.* out_format : pfs
 Date: Jan 31, 2002 12:57 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=framet_n2p.model -DEV=xlh
 -O=cn2.1/USPTO.spool/US08957709/runat_30012002.133258.9479/app_query.fasta_1.536
 -DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -FCAPOP=4.500 -QCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US08957709 -CGN1_1_59 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
 Query: US-08-957-709-70
 Query length: 471
 Database: A_Geneseq_1101.*
 Database sequences: 522463
 Database length: 74073290
 Search time (sec): 47.930000

score_list:	Strid Orig	zScore	EScore	Len	Documentation
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT:AAW72847				156	3.0e-91
/SID22/gcgdata/geneseq/geneseq/AA1993.DAT:AAW72847				173	2.8e-21
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AAW72847				189	2.7e-18
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AAW72847				160	1.4e-15
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				188	3.84.31
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				190	1.1e-13
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				190	8.5e-10
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				190	2.3e-08
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				177	6.0e-08
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				1143	8.2e-06
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				141	3.0e-05
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				164	3.1e-05
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				252	3.6e-05
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				181	3.7e-05
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				1105	6.9e-05
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				166	0.0017
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				198	0.0018
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				254	0.0030
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				1124	0.0210
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				1031	0.1307
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				149	0.0901
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				639	0.1478
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				1124	0.3170
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				154	0.2475
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				470	0.6404
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				142	0.4917
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				142	0.4917
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				201	0.6383
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				14	0.3959
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				154	0.8948
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				659	1.47
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				505	1.55
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				995	1.95
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				206	1.47.32
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				209	1.52
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				169	1.64
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				594	2.89
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				594	2.89
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847				860	3.28

alignment_scores:

/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AAW72847 + 74.00 122.71 4.18 1755
 /SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AAW72847 + 73.50 151.32 1.83 102
 /SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AAW72847 + 73.50 149.62 1.94 120
 /SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AAW72847 + 73.50 149.54 1.94 121
 /SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AAW72847 + 73.50 148.63 2.00 132

seq_name: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT:AAW72847

seq_documentation_block:

ID: AAW72847 standard; Protein; 156 AA.

XX AAW72847;

XX 01-MAR-1999 (first entry)

DE Polymerase enhancing factor P45 (dUTPase) component.

XX Polymerase enhancing factor; PEF; dUTPase; DNA polymerase; PCR;

KW amplification; sequencing; replication.

XX Pyrococcus furiosus strain DSM 3638.

XX Key Location/Qualifiers

FT Peptide 2..15

FT /note= "N-terminal peptide used to generate

XX primers"

XX WO9842860-A1.

XX 01-OCT-1998.

XX 20-MAR-1998; 98WO-US05497.

XX 24-OCT-1997; 97US-0957709.

XX 21-MAR-1997; 97US-0822774.

(STRA-) STRATAGENE.

Hansen CJ, Hogrefe H;

WPI; 1998-542284/46.

N-PSDB; AAV63860.

Polymerase enhancing factor proteins, extracts and complexes -

Improve the polymerisation activity of nucleic acid polymerases, for

use in amplification, sequencing and replication

Claim 17; Page 43; 161pp; English.

This is the amino acid sequence of the P45 component of the
 polymerase enhancing factor (PEF) of *Pyrococcus furiosus* DSM 3638.
 The sequence is predicted from a DNA sequence (see AAV63860) obtained
 from genomic DNA by PCR. P45 and P50 (see AAW72844) are the
 predominant components of PEF, which acts to enhance the activity of
 P. furiosus DNA polymerase. P45 functions as a dUTPase, and can be
 used to enhance nucleic acid replication, polymerisation or PCR
 reactions. The invention provides novel extracts, proteins and
 complexes that improve the polymerisation activity of nucleic acid
 polymerases, as well as DNA constructs and antibodies. Also
 included are methods for identifying compositions with PEF activity,
 for purifying and using these compositions, and specific extracts,
 proteins and complexes that function to enhance polymerase activity.
 Nucleic acid polymerase reactions can be enhanced (claimed) by
 mixing a nucleic acid template, at least 1 polymerase and a
 composition having polymerase enhancing activity. Kits are
 provided for replicating nucleic acids, for site-directed
 mutagenesis, for nucleic acid sequencing or amplification
 (preferably PCR or RT-PCR).

Sequence 156 AA;

Quality: 806.00 Length: 156
 Ratio: 5.167 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-08-957-709-70 x AAW2847 ..
 Align seg 1/1 to: AAW2847 from: 1 to: 156
 1 ATGCTACTCTCCAGCTGGAAATCAGAAAGAAATACTTATAGAGCCATT 50
 1 MetLeuLeuProAspTrpIysIleArgIysGluIleLeuIleGluProPh 17
 51 TTCTGAAGAAATCCCTCCAAACAGCAGGTTATCACCTCAGAGTGGCAGAG 100
 51 eSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg 34
 17 eSerGluGluSerLeuGlnProAlaGlyTyrAspLeuArgValGlyArg 34
 101 AGGCTTTTGTAAAGGGAAATTAATCGAGCTGGAAGGAAGGAAGAACTC 150
 34 LuAlaPheValIysGlyIysLeuIleAspValGluIysGluGlyVal 50
 151 GTTATCTCTCCAGGGAATACGCTTATCTTAACCTCGAGAGGATAAA 200
 51 ValIleProProArgIleTyrAlaLeuIleLeuThrLeuGluArgIle 67
 201 GTTCCCGCAGCATGTATGGGGGATATGAAGATAAGGACGACTTTAGCAA 250
 67 sLeuProAspValMetGlyAspMetIysIleArgSerLeuAla 84
 251 GAGAAGGGGTATTGGTCTCTTTTGGTGGTGGTGGTGGTGGTGGTGGT 300
 84 rgGluGlyValIleGlySerPheAlaTrpValAspProGlyTrpAspGly 100
 301 AACTTAAACACTAATGCTCTACAAATGCCTCAAAATGAACCTGTGCAATTAG 350
 101 AsnLeuThrLeuMetLeuTyrAsnAlaSerAsnGluProValGluLeu 117
 351 ATATGGAGAGATTTGTGCAGATCGCATTTATTAAGGCTAGAGGCTCGG 400
 117 gTyrGlyGluArgPheValGlnIleAlaPheIleArgLeuGluGlyPro 134
 401 CAAGAAACCTTACAGAGAACTATCAGGGCAGCAGCAGGTTAGCGTTT 450
 134 IaArgAsnProTyrArgGlyAsnTyrGlnGlySerThrArgLeuAlaPhe 150
 451 TCAAGAGAGAAAGAACTC 468
 151 SerLysArgLysLysLeu 156
 seq_name: /SIDS2/gcgdata/geneseq/AA1993.DAT: AAR51078
 seq_documentation_block:
 ID AAR51078 standard; Protein: 173 AA.
 AC AAR51078;
 DT 27-MAY-1994 (first entry)
 DE Polypeptide encoded by pDAM-L3 plasmid fragment.
 XX Ligase, thermostable; Desulfohalobium ambivalens;
 KW Archaeobacteria; detection; ligation; Atp; adenosine triphosphate;
 XX DE4217134-A.
 XX 25-NOV-1993.
 XX 23-MAY-1992; 92DE-4217134.
 XX 23-MAY-1992; 92DE-4217134.
 XX (BOE) BOEHRINGER MANNHEIM GMBH.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Jarsch M, Kaletta C, Kessler C, Kletzin A, Rueger R;
 WPI; 1993-378402/48.
 DR N-PSDB; AAG2300.
 XX Thermostable ligase from archaeobacteria - and DNA coding for it,
 PT useful for nucleic acid detection
 XX Example 3; Figure 1; 26pp; German.
 XX A sequence encoding a thermostable DNA ligase was cloned into the
 CC plasmid pDAM-L3. The DNA ligase is ATP dependent and is useful for
 CC detecting nucleic acids by hybridising two oligonucleotides with
 CC adjacent sequences of the target nucleic acid, ligating the two
 CC oligonucleotides (using the ligase) and detecting the ligation
 CC product. Analysis of the fragment revealed three other open reading
 CC frames (ORF's). This is a polypeptide encoded by ORF3
 XX Sequence 173 AA;
 SQ
 alignment_scores:
 Quality: 242.00 Length: 156
 Ratio: 2.283 Gaps: 3
 Percent Similarity: 67.949 Percent Identity: 35.897
 alignment_block:
 US-08-957-709-70 x AAR51078 ..
 Align seg 1/1 to: AAR51078 from: 1 to: 173
 34 ATACTTATAGAGCCATTCTCAAGATCGCTCCCAACCCAGCAGGTTATGA 83
 17 IleValIleSerProLeuThrGlnAspThrIleArgGluAsnGlyValas 33
 84 CTCAGAGTGGGC.....AGAGAGGCTT 106
 33 pLeuArgValGlyGlyGluIleAlaArgPheLysLysThrAspGluIleT 50
 107 TTGTTAAGGGGAAA.....TTAATCGACGTGGAAGCAAGA 144
 50 yrGluAspGlyLysAspProArgSerPheTyrGluIleGluLysGlyAsp 66
 145 AAGTCGTATTCTCCCAAGGAATACGCTTAATCTTAACCTCGAGAG 194
 67 GluPheIleIleTyrProAsnGluHisValLeuLeuValThrGluGlu 83
 195 GATAAGTTGCCGACGATGTATGGGGATATGAAGATAAGGACGAGTT 244
 83 rValLysLeuProAsnAspValMetAlaPheValAsnLeuArgSerSerP 100
 245 TAGCAAGAGAGGGGTATTGTTCTTTTGGTGGTGGTGGTGGTGGTGG 294
 100 heAlaArgLeuGlyLeuPheValProProThrIleValAspAlaGlyPhe 116
 295 GATGAAACTTAACACTAATGCTCTACAAATGCCTCAAAATGAACCTGTG 344
 117 GluGlyGlnLeuThrIleGluValLeu...GlySerAlaPheProVally 132
 345 ATTAAGATATGAGAGAGATTGTGCAGATCGCATTTATTAAGGCTAGAG 394
 132 sileLysArgGlyThrArgPheLeuHisLeuIlePheAlaArgThrLeu 149
 395 GTCCGGCAAGAAACCTTACAGAGAAACTATCAGGGGAGCACAAGTTA 444
 149 hrProValGluAsnProTyrHisGlyLysTyrGlnGlyGlnGlyVal 165
 445 GCGTTTTCAGAGAGAAAG 462
 166 ThrLeuProLysPheLys 171
 seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT: AAG92886

seq_documentation_block:
ID AAG92886 standard; Protein; 189 AA.
XX
AC AAG92886;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6640.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
XX
PR 07-APR-2000; 2000JP-0159162.
XX
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mitoquchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
DR N-PSDB; AAH68105.
XX
PT Novel polynucleotides derived from Corynebacterium, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
PS Claim 17: SEQ ID NO: 6640; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 189 AA;

alignment_scores:
Quality: 218.00 Length: 169
Ratio: 2.057 Gaps: 4
Percent Similarity: 62.722 Percent Identity: 33.728
alignment_block:
US-08-957-709-70 x AAG92886 ..
Align seg 1/1 to: AAG92886 from: 1 to: 189
1 ATGCTACTTCCAGACTGGAAATCAGAAAA.....GAAAT 35
1 MetLeuSerAspArgAspIleArgLysSerIleAspAlaGlyAspLe 17
36 ACTTATAGACCATTTCTGAAGAATCGCTCCAAACGACGAGTTATGACC 85
; |||||||||||: || :|||:|||||: |||

17 uGlyIleGluProPheAspAlaGluLeuIleGlnProSerSerValAspV 34
86 TCAGAGTGGGCGA.....GAGGCTTTTGTAAAGGGAAA..... 120
34 aIArgMetAspArgTyrPheArgValPheAsnAsnSerLysTyrThrHis 50
121TTAATCGAGCTGGA 134
51 IleAspProLysLeuAsnGlnAspGluLeuThrSerLeuValGluValG 67
135 AAGGAAGAAAGTCGTTATTCCTCCAAAGGAATACGCTTAAATCCTAA 184
67 uAspGlyGluGlyPheValLeuHisProGlyGluPheValLeuAlaSer 84
185 CCTCGAGAGGATAAAGTTGCCCGACGATGTTATGGGGATATGAAGATA 234
84 hrLeuGluLysPheThrLeuProAlaHisLeuAlaGlyArgLeuGluGly 100
235 AGGAGCAGTTTACCAAGAGAGGGGTTATT.....GGTTCTTTTGGCTTG 278
101 LysSerSerLeuGlyArgLeuGlyLeuLeuThrHisSerThrAlaGlyPh 117
279 GGTGACCCAGGATGGATGGAACTTAACATAATGCTCTACAATGCCCT 328
117 eIleAspProGlyPheSerGlyTyrIleThrLeuGluLeuSerAsnVala 134
329 CAATGAACCTCTCGAATTAAGATATGGAGAGAGATTTGTGCAGATCGCA 378
134 IaAsnLeuProIleThrLeuTrpProGlyMetLysValGlyGlnLeuAla 150
379 TTTATAAGGCTAGAGGTCGGCAAGAAACCTTACAGAGAGAAACTATCA 428
151 LeuPheGlnMetSerSerProAlaGluThrProTyrGlySerGlyLysLe 167
429 GGGGAGC 435
167 uGlySer 169
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB80139
seq_documentation_block:
ID AAB80139 standard; Protein; 160 AA.
XX
AC AAB80139;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:1012.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00923.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032140.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032237.
 PR 09-JUL-1999; 99DE-1032258.
 PR 09-JUL-1999; 99DE-1032259.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 PA (BADI) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

WPI: 2001-137957/14.

N-PSDB; AAF72258.

Nucleic acids from *Corynebacterium glutamicum* encoding metabolic pathway proteins, useful for producing fine chemicals in microorganisms, including organic acids, nonproteinogenic amino acids, and purine and pyrimidine bases -

Claim 20: Page 1554; 1737pp; English.

CC AAF71753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum MP nucleic acids are useful for the production of fine chemicals in microorganisms, including organic acids, nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, polyketides and enzymes.

XX Sequence 160 AA;

alignment_scores:
 Quality: 196.00 Length: 158
 Ratio: 2.000 Gaps: 4
 Percent Similarity: 62.025 Percent Identity: 32.911
 alignment_block:
 US-08-957-709-70 x AAB80139
 Align seg 1/1 to: AAB80139 from: 1 to: 160
 1 ATGCTACTCCGACTGGAAATCAGAAA.....GAAAT 35
 1 ValLeuLeuSerAspArgAspLeuGlySerIleAspAlaGlyAspLeu 17
 36 ACTTATAGACCATTTTCTGAAGAAATCGCTCCACACGACGAGTTATGACC 85
 17 uGlyIleGluProPheAspAlaGluIleGlnProSerSerValAspV 34
 86 TCAGAGTGGGCAGA.....GAGGCTTTTGTAAAGGGGAAA..... 120
 34 alArgMetAspArgTyrPheArgValPheAsnAsnSerLysTyrThrHis 50
 121TTAATCGACGTGGA 134
 51 IleAspProLysLeuAsnGlnAspGluLeuThrSerLeuValGluValGl 67
 135 AAAGGAAGGAAAAGTCGTTATTCCTCAAGGGAATACGCCTTAATCCTAA 184
 67 uAspGlyGluGlyPheValLeuHisProGlyGluPheValLeuAlaSerT 84
 185 CCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGATATGAAGATA 234
 84 hrLeuGluLysPheThrLeuProAlaHisLeuAlaGlyArgLeuGluGly 100
 235 AGCAGCAGTTTAGCAAGAGAGGGGTATT.....GGTTCTTTTGTCTTG 278
 101 LysSerSerLeuGlyArgLeuGlyLeuLeuThrHisSerThrAlaGlyPh 117
 279 GGTTCACCCAGGATGGGAACTTAACACTAATGCTCTACAATGCCT 328
 117 eileAspProGlyPheSerGlyTyrIleThrLeuGluLeuSerAsnVal 134
 329 CAAATGAACCTGCTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCA 378
 134 laAsnLeuProIleThrLeuTyrProGlyMetLysValGlyGlnLeuAla 150
 379 TTTATAAGCTAGAGGTCGCGCA 402
 151 LeuPheGlnMetSerSerProAla 158

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW89818

seq_documentation_block:

ID_ AAW89818 standard; Protein: 188 AA.

XX AC AAW89818;

XX AC AAW89818;

DT 18-FEB-1999 (first entry)

XX DE Protein encoded by clone B17 ORF4.

XX KW Antigen; immunogenic cluster family; vaccine; gastritis; diagnosis;

XX KW peptic ulcer; gastric adenocarcinoma; gastric lymphoma.

XX OS Helicobacter pylori.

XX XX WO9849314-A2.

XX PN 05-NOV-1998.

XX PD 27-APR-1998; 98WO-US08487.

XX PF 27-APR-1998; 98WO-US08487.

XX XX

155 GlyIleAlaGlnValLeuPhePheGluGlyAspAlaAlaCysAspValSe 171

411 TTAC.....AGAGGAACATATCAG 429

111 ::::|||||

171 rTyRAlaGluArgGlnGlyLysTyrGln 180

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY34994

seq_documentation_block:

ID AAY34994 standard; Protein; 190 AA.

XX AC AAY34994;

XX DT 13-SEP-1999 (first entry)

XX DE Protein involved in intermediate metabolism of nucleotides.

XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.

XX OS Chlamydia pneumoniae.

XX PN WO9927105-A2.

XX PD 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-IB01890.

XX PR 04-NOV-1998; 98US-0107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX PA (GEST) GENSET.

XX PI Griffais R;

XX DR WPI; 1999-357842/30.

XX PT Genome sequence of Chlamydia pneumoniae

XX PS Page 901; Disclosure; 1912pp; English.

XX CC AAY34584-Y35879 represent the proteins encoded by all the open reading
frames in the complete genome (see AAY34584-Y35879) of Chlamydia pneumoniae.
XX CC C. pneumoniae causes respiratory disease such as pneumonia and
XX CC bronchitis and is thought to be a contributing factor in heart
XX CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX CC nodosum or pharyngitis. The polypeptides encoded by the open reading
frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
XX CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX CC nucleotide sequences can also be used as immunogenic compositions,
XX CC especially where the vector directs the expression of a neutralising
XX CC epitope of C. pneumoniae.

XX SQ Sequence 190 AA;

alignment_scores:

Quality: 138.00 Length: 164
Ratio: 1.484 Gaps: 5

Percent Similarity: 56.707 Percent Identity: 24.390

alignment_block:

US-08-957-709-70 x AAY34994 ..

Align seg 1/1 to: AAY34994 from: 1 to: 190

37 CTTATAGACCATTT.....TCTGAAGATCG... 63
:|||||:|||||

18 MetIleHisProPheValAsnGlyGlnValAsnValAsnGluGluThrG1 34
:|||||:|||||

64CTCCACACGACGGTTATGACCTCAGAG 91
|||||: |||||||

34 yGluLysLeuIleSerTyrGlyLeuSerSerTyrGlyTyrAspLeuArgL 51

92 TGGCGAGAGAG..... 102

:|||||

51 euSerArgGluPheLysValPheThrAsnValTyrAsnSerValValasp 67

103GCTTTTGAAGGGAAATTAATCGAGTGGAAAGAGAGAAA 146

:|||||:|||||

68 ProLysCysPheThrGluAspIleSerIle...ThrAspAspVa 83

147 ACTCGTTATTCTCCAAGGAATACGCTTAATCTTAACCTCGAGAGGA 196

:|||||:|||||

83 IcySileValProProAsnSerPheAlaLeuAlaArgSerValGluTyrP 100

197 TAAAGTTCCCGACGATGTTATGGGGGATATGAAGATAAGGACGAGTTTA 246

:|||||:|||||

100 heArgIleProArgAsnValLeuThrMetCysIleGlyLysSerThrTyr 116

247 CGAAGAGAGGGTTATTGGTCTTTTCTGGTTGACCCAGGATGGGA 296

:|||||:|||||

117 AlaArgCysGlyIleIleValAsnValThrProPheGluProGluTyrpG1 133

297 TGGAAACTTAACACTAATGCTCTACAATGCCCTCAATGAACCTGCGAAT 346

:|||||:|||||

133 uGlyHisValThrIleGluIleSerAsnThrThrProLeuProAlaLysI 150

347 TAAGATATGAGAGAGATTGTGCAGATCGCATTTATAAGGCTAGAGGCT 396

:|||||:|||||

150 IeTyrAlaAsnGluGlyIleAlaGlnValLeuPhePheGluSerSerThr 166

397 CCGCAAGAAACCCCTTAC.....AGAGAAACTATCATCAG 429

:|||||:|||||

167 ThrCysGluValSerTyrAlaAspArgLysGlyLysTyrGln 180

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY29169

seq_documentation_block:

ID AAY29169 standard; Protein; 177 AA.

XX AC AAY29169;

XX DT 25-OCT-1999 (first entry)

DE DE Amino acid sequence of a virulence factor encoded by ORF19027c.

XX KW Human pathogen; virulence polypeptide; virulence factor;

XX KW pathogenic infection; Pseudomonas aeruginosa infection.

XX OS Pseudomonas aeruginosa.

XX PN WO9927129-A1.

XX PD 03-JUN-1999.

XX PF 25-NOV-1998; 98WO-US25247.

XX PR 25-NOV-1997; 97US-0066517.

XX PA (GEHO) GEN HOSPITAL CORP.

XX PI Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;

XX PI Rahme LG, Tan M, Tsongalis J;

XX DR WPI; 1999-357851/30.

XX PT Virulence factors useful in developing disease treatments

XX PS Disclosure; Fig 3; 228pp; English.

XX CC The present sequence represents a Pseudomonas aeruginosa polypeptide
sequence. P. aeruginosa is an opportunistic human pathogen present in
XX CC soil water and plants. The specification describes virulence polypeptides
XX CC and nucleic acid sequence encoding such polypeptides. These sequences

130 SerAlaArgAlaAlaGlyTyrAspLeuTyrSerAlaTyrAsp..... 143
111 TAAGGGAAATTAATCGAGTGGAAAGAAAGTGGTATTCCTC 160
144TyrThrIleProp 148
161 CAAGGGAATACGCCCTTAATCAACCTCGAGAGGATAAAGTTGCCGAC 210
148 roMetGluLysAlaValVallyThrAspIleGluIleAlaLeuProSer 164
211 GATGTTATGGGGATATGAGATAGGACGAGTTAGCAAGAAGGGGT 260
165 GlyCysTyrGlyArgValAlaProArgSerGlyLeuAlaAlaLysHisPh 181
261 TATTCGTTCTTTTCTGGT...GACCCAGGATGGGATGCAACTAA 307
181 elleAspValGlyAlaGlyValIleAspGluAspTyrArgGlyAsnValG 198
308 CACTAATGCTCTACAACTCAATCAACCTGTCGAATTAAGATATGGA 357
198 lyValValLeuPheAsnGlyLysGluLysPheGluValLysGly 214
358 GAGAGATTGTCAGATCGCATTTATAAGGCTA.....GAGGGTCC 398
215 AspArgIleAlaGlnLeuIleCysGluArgIlePheTyrProGlu.IleG 231
399 GCAAGAAACCTTACAGAGGAACATCAGGGGAGCACACAGGT 442
231 luGluValGlnAlaLeuAspThrGluArgGlySerGlyGly 245

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA29165
seq_documentation_block:
ID: AAY29165 standard; Protein; 181 AA.
XX AAY29165;
XX
XX
XX 25-OCT-1999 (first entry)
XX
XX Amino acid sequence of a virulence factor encoded by ORF17793c.
XX
XX Human pathogen; virulence polypeptide; virulence factor;
XX pathogenic infection; Pseudomonas aeruginosa infection.
XX
XX Pseudomonas aeruginosa.
XX
XX W09927129-A1.
XX
XX 03-JUN-1999.
XX
XX 25-NOV-1998; 98WO-US25247.
XX
XX 25-NOV-1997; 97US-0066517.
XX
XX (GEO) GEN HOSPITAL CORP.
XX
XX Ausubel F, Cao H, Drenkard E, Goodman HW, Mahajan-Miklos S;
XX Rahme LG, Tan M, Tsongalis J;
XX
XX WPI; 1999-357851/30.
XX
XX Virulence factors useful in developing disease treatments
XX
XX Disclosure; Fig 4; 228pp; English.

The present sequence represents a Pseudomonas aeruginosa polypeptide sequence. P. aeruginosa is an opportunistic human pathogen present in soil water and plants. The specification describes virulence polypeptides and nucleic acid sequence encoding such polypeptides. These sequences can be used to identify a compound which is capable of decreasing the expression of a pathogenic virulence factor. Compounds that inhibit the expression or activity of virulence factor polypeptides can be used to treat pathogenic infections, especially where the infection

CC is a P. aeruginosa infection.
CC note: the sequences given in the specification were poorly legible, and
CC in some instances assumptions were made as to the identity of the
CC residue; it is therefore possible that the sequence given below is
CC not entirely correct.
XX

SQ Sequence 181 AA;

alignment_scores:
Quality: 112.00 Length: 173
Ratio: 1.217 Gaps: 5
Percent Similarity: 53.179 Percent Identity: 24.855

alignment_block:
US-08-957-709-70 x AAY29165
Align seg 1/1 to: AAY29165 from: 1 to: 181
4 CTACTTCCAGACTGGAAAATCAGAAAAGAAATATATAGAGCCATTTTC 53
13 LeuValAlaGlnCysAlaIleCysGluLysThrLeuPheAspGluPheSe 29
54 T.....GAGAATCGCTCCCAACACG 73
29 rLeuLysIleGlnLeuGlyHisThrTyrTyrGluProLysSerLeuProA 46
74 CAGGTTATGACCTCAGTGGCGCAGAGGCT.....TTT 108
46 laSerAlaSerIleValTyrGlySerHisProAlaProSerThrPhePhe 62
109 GTAAGGGGAAATTAATCGACGTGGAAAGAAAGGAAAGTCTTTATTC 158
63 LeuGluProLys.....GluIleGlnGlnAsnLeuValLeuLy 75
159 TCCAAGGGAATACGCTTAATCCTTAACCTCGAGAGGATAAAGTTGCCGG 208
75 sSerGlyGluGlnValIleThrCysSerLysHisArgTyrLysIleProL 92
209 AGCATGTTATGGGGATATGAAGATAAGGACGATTAGCAAGA..... 252
92 euAspTyrPheGlyLeuValGlnThrLysGlyThrLeuAlaArgLeuPhe 108
253 GAAGGGTTATTGGTTCTTTGTTGGTTGACCCAGGATGGATGGAAA 302
109 ValGlnValThrCysAsnAspGlyGlnValGluProGlyPheAspGlyTy 125
303 CTTAACTACTATGCTCTACATGCCTCAATGAACCTGTCTGAATTAAGAT 352
125 rValThrLeuGluIleValAsnMetSerProThrThrIleGluIleProA 142
353 ATGGAGAGAGATTGTGTCAGATCGCATTTATAAGGCTAGAGGTCGGCA 402
142 laValSerAspIleAlaGlnLeuTyrLeuValLysCysSerThrSerAla 158
403 AGAAACCTTACAGAGAGAACTAT.....CAGGGGAGCAC 437
159 SerGluProTyrHisGlyArgTyrMetAspAlaAlaLysLysGlyProPh 175
438 AAGGTTAGCGTTTTCANAG 456
175 rLeuAlaValPheArgLys 181

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT:AA291667
seq_documentation_block:
ID: AAP71667 standard; Protein; 1105 AA.
XX
XX AAP71667;
XX
XX 26-APR-1991 (first entry)
XX
XX Translation product of pol transcript of VISNA sheep lentivirus.
XX

OM of: US-08-957-709-70 to: SPTREMBL_17:* out_format : pfs

Date: Jan 31, 2002 1:05 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet-n2p.model -DEV=xlh
-O=/cn2_1/USPTO.spool/US08957709/runat_30012002.133259_9563/app_query.fasta_1.536
-DB=SPTREMBL_17 -QFWT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blowsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US08957709.ecgn1_114 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-08-957-709-70

Query length: 471

Database: SPTREMBL_17:*

Database sequences: 473505

Database length: 146272329

Search time (sec): 66.180000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_bacteria:Q9KFV3	+	242.50	376.43	6.1e-13	177	Q9KFV3 bacillus halodurans. deo
sp_virus:071028	+	237.00	368.82	1.8e-12	158	071028 archaeal virus sirv. dut
sp_bacteria:Q9CB17	+	197.50	306.18	4.6e-09	190	Q9CB17 mycobacterium leprae. ph
sp_bacteria:Q9PN07	+	189.50	293.96	2.3e-08	186	Q9PN07 campylobacter jejuni. pc
sp_archaea:Q9HSG3	+	184.50	285.84	6.2e-08	195	Q9HSG3 halobacterium sp. (stra)
sp_bacteria:Q9HYC9	+	172.00	266.77	7.4e-07	188	Q9HYC9 pseudomonas aeruginosa.
sp_bacteria:Q9PP86	+	164.50	255.03	3.3e-06	191	Q9PP86 xylella fastidiosa. deox
sp_archaea:Q9VG32	+	159.00	247.78	9.7e-05	163	Q9VG32 aeropyrum pernix. 163aa
sp_bacteria:Q9JRE8	+	155.00	240.45	2.2e-05	188	Q9JRE8 neisseria meningitidis (
sp_archaea:Q27642	+	153.00	239.15	3.2e-05	150	Q27642 methanobacterium thermo
sp_bacteria:Q774073	+	135.50	210.04	0.0010	193	P74073 synechocystis sp. (stra)
sp_virus:Q9EG61	+	126.50	181.93	0.0068	1146	Q9EG61 equine infectious anem
sp_archaea:Q29157	+	126.00	196.43	0.0068	168	Q29157 archaeoglobus fulgidus.
sp_virus:Q9WHF3	+	125.50	195.57	0.0075	170	Q9WHF3 agrotis segetum granulos
sp_virus:Q9W706	+	125.50	180.44	0.0078	1138	Q9W706 equine infectious anem
sp_virus:Q89468	+	125.50	180.38	0.0078	1146	Q89468 equine infectious anem
sp_virus:Q89472	+	125.50	180.38	0.0078	1146	Q89472 equine infectious anem
sp_virus:Q66733	+	123.50	185.23	0.0114	422	Q66733 equine infectious anem
sp_bacteria:Q9EP46	+	123.50	177.27	0.0117	1148	Q9EP46 equine infectious anem
sp_bacteria:Q9RWF1	+	123.00	190.44	0.0124	199	Q9RWF1 zymomonas mobilis. deox
sp_virus:Q992J6	+	120.50	172.72	0.0211	1134	Q992J6 equine infectious anem
sp_virus:Q9EP41	+	119.50	171.08	0.0258	1148	Q9EP41 equine infectious anem
sp_invertebrate:Q9V311	+	119.00	184.70	0.0275	188	Q9V311 drosophila melanogaste
sp_archaea:Q58502	+	118.50	185.16	0.0302	161	Q58502 methanococcus jannaschi
sp_virus:Q992K2	+	117.50	168.08	0.0384	1134	Q992K2 equine infectious anem
sp_virus:Q9DKV8	+	116.50	166.72	0.0468	1107	Q9DKV8 caprine arthritis encep
sp_archaea:Q9HMF3	+	103.00	160.96	0.6569	165	Q9HMF3 halobacterium sp. (stra)
sp_bacteria:Q9A1K2	+	102.00	160.28	0.7995	148	Q9A1K2 streptococcus pyogenes.
sp_rodent:Q9JJ44	+	101.00	158.01	0.9769	162	Q9JJ44 mus musculus (mouse). du
sp_rodent:Q9CQ43	+	101.00	158.01	0.9769	162	Q9CQ43 mus musculus (mouse). du
sp_rodent:Q9CU90	+	101.00	156.17	0.9816	204	Q9CU90 mus musculus (mouse). 51
sp_virus:Q72165	+	100.50	157.33	1.08	160	Q72165 orf virus. dutpase homol
sp_virus:Q9WMY7	+	99.00	147.14	1.48	430	Q9WMY7 dioscorea alata bacilli
sp_virus:Q98W23	+	99.00	137.17	1.48	430	Q98W23 dioscorea alata bacilli
sp_virus:Q84809	+	99.00	139.77	1.51	1086	Q84809 puma lentivirus 14. gad
sp_plant:Q9STG6	+	98.50	153.94	1.61	166	Q9STG6 arabidopsis thaliana (md
sp_virus:P87630	+	95.00	149.49	3.21	147	P87630 cowpox virus (cpv). 41kb
sp_virus:Q9JG65	+	93.50	147.28	4.32	145	Q9JG65 fowlpox virus. (fpv) 3
sp_virus:Q9J880	+	93.00	146.61	4.77	143	Q9J880 spodoptera exigua nucle
sp_virus:Q9Q8S9	+	90.50	142.47	7.85	148	Q9Q8S9 myxoma virus (strain lau

sp_virus:Q9JFF1 + 89.50 140.97 9.57 147 ! Q9JFF1 vaccinia virus (strai
sp_organelle:Q9G4C0 - 88.00 132.43 13.10 321 ! Q9G4C0 thraustochytrium aure
sp_virus:Q66933 + 87.50 121.70 14.84 1123 ! Q66933 feline immunodeficie
sp_virus:P90246 + 87.50 121.51 14.85 1150 ! P90246 feline immunodeficie
sp_bacteria:Q9A253 + 87.00 136.68 15.74 155 ! Q9A253 caulobacter crescentu

seq_name: sp_bacteria:Q9KFV3

seq_documentation_block:

ID Q9KFV3 PRELIMINARY; PRT; 177 AA.

AC Q9KFV3;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.

GN BH0368.

OS Bacillus halodurans.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=86665;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C-125 / JCM 9153;

RX MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,

Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

Horikoshi K.;

RA "Complete genome sequence of the alkaliphilic bacterium Bacillus

RT halodurans and genomic sequence comparison with Bacillus subtilis.";

RL Nucleic Acids Res. 28:4317-4331(2000).

DR EMBL; AP001508; BAB04087.1;

DR InterPro; IPR003232; dCTP_deaminase.

DR Pfam; PF006592; dutpase; 1.

DR ProDom; PD004900; dCTP_deaminase; 1.

KW Complete proteome.

SQ SEQUENCE 177 AA; 19897 MW; D15AE75387847E2B CRC64;

alignment_scores:

Quality: 242.50 Length: 156

Ratio: 2.310 Gaps: 6

Percent Similarity: 67.308 Percent Identity: 37.179

alignment_block:

US-08-957-709-70 x Q9KFV3 ..

Align seg 1/1 to: Q9KFV3 from: 1 to: 177

28 AAAGAAATACCTATAGAGCCATTTCTGAAGAAATCGCTCCACACGAGG 77

15 LysGluLeuGluIleThrProLeuThrGluGluGlnIleGlnProAla 31

78 TTATGACCTCAGAGTGGCAGAGAGCGCTTTTGTGTT.....A 112

31 rValAspLeuArgLeuGlyProHis....PheValThrIleAspSerL 47

113 AGGGGAATTAATCGACCTGGAAAG..... 138

47 yGluAlaValIleSerPheGluArgProIleArgTyrArgGluTyrThr 63

139GAAGAAATCGTTTATTCCTCCAGGAGTACGCTTAATCCT 182

64 ThrSerSerGluThrIleValLeuProProHisThrPheLeuLeuAlaTh 80

183 AACCTCGAGAGGATAAAGTTGCCGAGATGTTATGGGGATATGAGA 232

80 rThrMetGluThrValLysLeuProAsnHisLeuThrAlaPheValGlu 97

233 TAAGGACGAGTTTAGCAAGAGAAGGGTT...ATTGGTTCCTTTGCTGG 279

97 lyArgSerSerValGlyArgLeuGlyLeuPheIleGlnAsnAlaGlyTyr 113

```

280 GTTGACCCAGGATGGATGGAACCTTAACACTAATGCTCTACAATGCCTC 329
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 ValAspProGlyPheAsnGlyGlnIleThrLeuGluLeuPheAsnAlaAs 130
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 AAATGAACCTGCTCGAATTAAGATATGAGAGAGATTTCGAGATCGCAT 379
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 nArgLeuProIleGluLeuProIleGlyArgGlyArgIleCysGlnLeuValp 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
380 TTATGAAGCTAGAGGTCGCGCAGAAACCCCTACAGAGGAACCTAT... 426
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 heAlaGluValThrGlyGluValAla...ProTyrGlnGlyLysTyrLeu 162
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
427 .....CAGGGGAGCACA 438
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 PheGlnLysGlyAlaThr 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: sp_virus:071028
seq_documentation_block:
ID 071028 PRELIMINARY; PRT; 158 AA.
AC 071028;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DUTPASE (EC 3.6.1.23) (DUTP PYROPHOSPHATASE) (DEOXYRIDINE-
DE TRIPHOSPHATASE).
OS Archaeal virus SIRV.
OC Viruses; unclassified viruses.
OX NCBI_TaxID=66287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165772; PubMed=9497317;
RA Prangishvilli D., Klenk H.P., Jakobs G., Schmiedchen A., Hanselmann C.,
RA Holz I., Zillig W.;
RT "Biochemical and phylogenetic characterization of the dUTPase from the
RT archaeal virus SIRV.";
RL J. Biol. Chem. 273:6024-6029(1998).
CC -/- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
DR EMBL; AF022221; AAC15873.1; -.
DR InterPro; IPR001428; dUTPase.
DR ProDom; PRD003232; dUTPase.
DR Pfam; PF00692; dUTPase.
DR ProDom; PD004900; dUTP_deaminase; 1.
KW Hydrolase.
SQ SEQUENCE 158 AA; 17906 MW; 5BA47765E0889190 CRC64;

alignment_scores:
Quality: 237.00 Length: 157
Ratio: 2.155 Gaps: 2
Percent Similarity: 70.064 Percent Identity: 34.395

alignment_block:
US-08-957-709-70 x 071028 ..
Align seg 1/1 to: 071028 from: 1 to: 158

1 ATGCTACTTCAGACTGGAATAATCAGA.....AAAGAAAT 35
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetIleuSerAspArgAspIleuysThrTyrIleAsnSerLysLysLe 17

36 ACTTATAGAGCCATTTCTGAAGATCGCTCCACACGAGGTTATGACC 85
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 ualIleAsnProIleuSerGluAspThrIleArgIleuAsnGlyValAspL 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 TCAGAGTGGGAGAGAGCGCTTTTGTAAAGGGAAATTAATCGACGTGAA 135
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 euLysIleGlyAsnGluIleValArgIleLysGluAsnMetGluLysGlu 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 AAGGAAGAAAGTCGTATTCTCCAGGAATACGCCTTAATCCCTAAC 185
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ValGlyAspGluPheIleIleTyrProAsnGluHisValLeuLeuThrTh 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

186 CCTGAGAGGATAAAGTTGCCGACGATGTTATGGGGATATGAAGATAA 235
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 rLysGluTyrIleLysLeuSerAsnAspIleIleAlaPheCysAsnLeuA 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
236 GGAGCAGTTTTCAGCAAGAGAGGGTTATTGCTTCTTTTGGTTGGTTCAC 285
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 rGserThrPheAlaArgLysGlyLeuLeuIleProThrIleValAsp 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
286 CCAGATGGGATGAAACTTAACATAATGCTCTACAAATGCTCAAAATGA 335
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 AlaGlyPheGluGlyGlnLeuThrIleGluLeuVal...GlySerSerI 116
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
336 ACCTGTCCAATTAAGATATGAGAGAGATTTCGAGATCGCATTTATAA 385
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 eProValLysLeuLysSerGlyGluArgPheLeuHisLeuIlePheAla 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
386 GGCTAGAGGTCGCGCAAGAACCTTACAGAGGAACCTATCAGGGGAGC 435
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
133 rGThrLeuThrProValGluLysProIleAsnGlyLysTyrGlnLysGln 149
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 ACAAGGTTAGCGTTTCAAG 456
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 LysGlyValThrLeuAlaLys 156
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: sp_bacteria:Q9CB17
seq_documentation_block:
ID Q9CB17 PRELIMINARY; PRT; 190 AA.
AC Q9CB17;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
GN ML2507.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL583925; CAC32024.1; -.
DR InterPro; IPR003232; dCTP_deaminase.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
KW Complete proteome.
SQ SEQUENCE 190 AA; 20820 MW; 1CA936700500B6FA CRC64;

alignment_scores:
Quality: 197.50 Length: 173
Ratio: 1.955 Gaps: 4
Percent Similarity: 58.382 Percent Identity: 30.636

alignment_block:
US-08-957-709-70 x Q9CB17 ..
Align seg 1/1 to: Q9CB17 from: 1 to: 190

1 ATGCTACTTCAGACTGGAATAATCAGAAAGAAATACTT..... 39

```

```

||||| 1 MetLeuLeuSerAspArgAspLeuArgAlaGluIleThrAlaGlyArgPh 17
40 .....ATAGACCCATTTCTCAGAAATCGCTCCCAACAGCAGGTTATGACC 85
17 eSerIleAspProPheAspThrLeuValGlnProSerSerIleAspV 34
86 TCAGAGTG..... 93
34 alArgLeuAspCysMetPheArgValPheAsnAsnThrArgTyrThrHis 50
94 .....GGCAGAGAGGCTTTGTTTAAAGGGGAAATTAATCGAGTGA 134
51 IleAspProAlaArgGlnGlnAspGluLeuThrSerLeuValGluLeuVa 67
135 AAAGGAAGGAAGTCTTATCTCTCAAGGGAATACGCTTAATCTAA 184
67 lAspGlyGluProPheValLeuHisProGlyGlyPheValLeuGlySerT 84
185 CCTCGAGAGGATAAAGTTCGCCGACGATGTTATGGGGGATATGAAGATA 234
84 hrLeuGluLeuPheThrLeuProGluAspLeuAlaGlyArgLeuGluGly 100
235 AGGACGAGTTTACAGAGAGGAGGTTATT.....GTTCTTTTGTCTG 278
101 LysSerSerLeuGlyArgLeuGlyLeuLeuThrHisSerThrAlaGlyPh 117
279 GGTGACCCAGGATGGATGGAATTAACACTTAATGCTCTACAAATCCCT 328
117 eIleAspProGlyPheCysGlyHisIleThrLeuGluLeuSerAsnVala 134
329 CAATGAACCTGTCGAATTAAGATATGAGAGAGATTGTGTCAGATCGCA 378
134 laAsnLeuProIleThrLeuThrProGlyMetIleGlyGlnLeuCys 150
379 TTTTATAGGCTAGAGGTCGGCAAGAAACCTTAC..... 414
151 ValLeuArgLeuThrSerProAlaGluHisProTyrGlySerAlaSerAl 167
415 .AGAGGAACATCAGGGG 432
167 aglySerLysTyrGlnGly 173

seq_name: sp_bacteria:Q9PN07

seq_documentation_block:
ID Q9PN07 PRELIMINARY; PRT; 186 AA.
AC Q9PN07;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE POSSIBLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13).
GN DCD OR CJI292.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.

seq_documentation_block:
ID Q9HSG3 PRELIMINARY; PRT; 195 AA.
AC Q9HSG3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
GN DTD OR VNG0245G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.

SQ SEQUENCE 186 AA; 20699 MW; 68DC660FB28FD33C CRC64;

alignment_scores:
Quality: 189.50 Length: 180
Ratio: 1.788 Gaps: 7
Percent similarity: 58.889 Percent Identity: 29.444

alignment_block:
US-08-957-709-70 x Q9PN07 ..

Align seq 1/1 to: Q9PN07 from: 1 to: 186

13 GACTGGAAATCAGAAAGAAATATT.....ATAGACCCATT 50
:||||| ||||||| |||
7 AsnTrp...IleArgLysMetAlaLeuGluHisLysMetIleGluProPh 22
51 TTCTGAAGATCG.....CTCCAACGAC 73
| ||| :|
22 eCysGluAlaAsnIleGlyLysGlyValValSerTyrGlyLeuSerSert 39
|||||
74 CAGGTTATGACCTCAGAGTGGCAGAGAG..... 102
|||||
39 YrGlyTyrAspIleArgValGlyArgGluPheLysIlePheThrAsnVal 55
103 .....GCTTTTGTAAAGGGGAAATTAATCGA 128
|||||
56 AsnSerThrValValAspProLysAsnPheValGluGluAsnValValas 72
129 CGTGGAAAGGAGGAAAGTC...GTTATTCCTCCAGGGAATACGCCT 175
| ||||| ||| :|||
72 pPhe.....GluGlyAspValCysIleValProAlaAsnSerPheAlaL 87
176 TAATCCTTAACCTCGAGAGGTAAGTTGCCGCGAGGATTTATGGGGAT 225
|| |||||
87 euAlaArgThrIleGluTyrPheLysMetProAspAsnValIleAlaIle 103
226 ATGAAGATAAGGACGAGTTTAGCAAGAGAGGGTTATTGTTCTTTTGC 275
:|||||
104 CysLeuGlyLysSerThrTyrAlaArgCysGlyIleIleValAsnValTh 120
276 TTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAATGCTCTACAATG 325
:|||||
120 rProPheGluProGlyPheGluGlyHisIleThrIleGluIleSerAsnt 137
326 CCTCAATGAACCTGTCGAATTAAGATATGAGAGAGATTGTTCGAGATC 375
:|||||
137 hrThrProLeuProAlaLysIleTyrAlaAsnGluGlyIleAlaGlnVal 153
376 GCATTTTATAGGCTAGAGGTCGGCAAGAAACCTTAC.....AG 416
|||||
154 LeuPheLeuGlnGlyAspGluLysCysAspThrThrLysAspLysLy 170
417 AGAAATATCAGGGGAGCAGACCAAGGTTAGCGTTTTCAAAAG 456
:|||||
170 sGlyLysTyrGlnAlaGlnThrGlyIleThrLeuProArg 183

seq_name: sp_archaea:Q9HSG3

seq_documentation_block:
ID Q9HSG3 PRELIMINARY; PRT; 195 AA.
AC Q9HSG3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
GN DTD OR VNG0245G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE-20504483; PubMed-11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithaus B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt T., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RA "Genome sequence of Halobacterium species NRC-1".
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AB004988; AAG18843.1; -.
DR InterPro: IPR003232; dUTPase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Complete proteome.
SQ SEQUENCE 195 AA; 21542 MW; 86CCDA4D48BC0C9D CRC64;

alignment_scores:
Quality: 184.50 Length: 173
Ratio: 1.724 Gaps: 7
Percent Similarity: 61.850 Percent Identity: 31.214

alignment_block:
US-08-957-709-70 x Q9HSG3
Align seg 1/1 to: Q9HSG3 from: 1 to: 195

1 ATGCTACTCCAGACTGGAAATCAGAAAA.....GAAAT 35
|||||
1 MetileLeuSerAspGlnAspIleLeuAlaArgLeuAlaAspGlyAspLe 17
36 ACTTATAGAGCCATTTCCTGAAGATCGCTC.....CAACGACGAGGTT 79
: |||||
17 ualallegluProLeuGluAspValAspLeuGlnValGlnProAlaSerV 34
80 ATGACCTCAGAGTGGGACAGAG.....GCTTTTGTAAAGGGAAATTA 123
|||||
34 alaspValArgLeuGlyArgArgPheLeuGluPheGluArgAlaAsnVal 50
124ATCGAC..... 129
51 ProCysIleHisProAsnArgGluAspGluValAspGluTyrValThrG1 67
130GTGAAAGGAGGAAAGTGGTATTCCTCCAGGGAATACG 172
|||||
67 uThrValValgluaspGlyaspGluPheIleLeuHisProGlyAspPheV 84
173 CTTAATCCTAACCTCGAGAGGATAAGTTGCCCGGAGGATGTTATGGG 222
|||||
84 alLeuGlyThrLysGluArgValGluValProArgAspLeuValAla 100
223 GATATGAAGATAAGAGAGGTTTACGAAGAGAGGGGTTATT.....GG 266
: |||||
101 GlnValGluGlyArgSerSerLeuGlyArgLeuAlaValValHisAl 117
267 TTTCTTTTCTGGTGTGACCCAGGATGGATGGAACTTAACACTAATGC 316
: |||||
117 aThrAlaGlyPheIleAspProGlyPheAsnGlyArgValThrLeuGluL 134
317 TCTACATCGCTCAATGAACCTGCTCAATTAAGATATGAGAGAGATTT 366
|| |||||
134 euSerAsnLeuGlyLysValProValAlaLeuThrProGluMetArgile 150
367 GTGCAGATCGCATTTATAAGGCTAGAGGTCGCGCAAGAAACCCCTACAG 416
151 SerGlnLeuValPheThrGluLeuThrSerProAlaAspArgProTyr... 166
417 AGAAACTATCAGGGGAGC 435
167 .GlyaspGluArgGlySer 172

seq_name: sp_bacteria:Q9HYC9
seq_documentation_block:
ID Q9HYC9 PRELIMINARY; PRT; 188 AA.
AC Q9HYC9
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
GN PA3480.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]_TaxID=287;
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy S., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AB004769; AAG06868.1; -.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Complete proteome.
SQ SEQUENCE 188 AA; 21154 MW; 9A811B5F17B2A996 CRC64;

alignment_scores:
Quality: 172.00 Length: 171
Ratio: 1.737 Gaps: 4
Percent Similarity: 57.895 Percent Identity: 26.316

alignment_block:
US-08-957-709-70 x Q9HYC9
Align seg 1/1 to: Q9HYC9 from: 1 to: 188

37 CTTATAGAGCCATTTCTGAAGATCGCTCCACACGCA..... 75
: |||||
18 MetileGluProPheValGluArgGlnValArgGlyAlaAspSerAr 34
76GGTTATGACCTCAGAGTGGGCA 97
|||||
34 gValleSerTyrGlyValSerSerTyrGlyTyrAspValArgCysAlaI 51
98 GAGAG..... 102
|||
51 laGluPheLysValPheThrAsnIleHisSerAlaValValAspProlys 67
103 GCTTTTGTAAAGGGGAAATTAATCGACGTGGAAAGGAAAGAAAGTCGT 152
|||
68 AsnPheAspGluLysSerPheValAspIleAsnSerAsp...ValCysII 83
153 TATTCCTCCAAGGAATACGCTTAACTTAACTCCCTCAGAGAGGATAAAGT 202
: |||||
83 eileProAsnSerPheAlaLeuAlaArgThrValGluTyrPheArgI 100
203 TGCCCGACGATGTTATGGGGATATGAAGATGAAGAGCAGCTTTAGCAAGA 252
: |||||
100 leProArgAspValLeuThrIleCysLeuGlyLysSerThrTyrAlaArg 116
253 GAAGGGGTTATGCTTCTTTGCTTGGTTGACCCAGGATGGGATGAAA 302
|||||

117 CysGlyllelleValAsnValThrProLeuGluProGluTrpGluGlyHi 133
 303 CTTAACACTAATGCTACAAATGCTCAAAATGACCTGTCGAATTAAGAT 352
 ::::||||| :::: |||:||||| :|||:||||| :|||:||||| :|||:|||||
 133 sValThrLeuGluPheSerAsnThrThrAsnLeuProAlaLysileTyra 150
 353 ATGAGAGAGATTGTCGACATCCCATTTATTAAGCGTAGAGGTCGGCA 402
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 150 laAsnGluGlyValAlaGlnMetLeuPheLeuGlnSerAspGluAlaCys 166
 403 AGAAACCTTACAGA.....GGAACTATCAGGGGACCAAGGTT 443
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 167 GluValSerTyrlsAspArgGlyGlyLysTyrlGlnGlyGlnArgGlyVa 183
 444 AGCGTTTTCAAAG 456
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 183 lThrLeuProLys 187

seq_name: sp_bacteria:Q9PFB6

seq_documentation_block:
 ID Q9PFB6 PRELIMINARY; PRT; 191 AA.
 AC Q9PFB6;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
 GN XF0762.

OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.V., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 DR EMBL: AF003917; AAF83572.1;
 DR InterPro; IPR003332; dCTP_deaminase.
 DR Prodom; PD004900; dCTP_deaminase; 1.
 KW Complete proteome.
 SQ SEQUENCE 191 AA; 21531 MW; D7B23653F94B3649 CRC64;

alignment_scores:
 Quality: 164.50 Length: 141

Ratio: 1.891 Gaps: 4
 Percent Similarity: 61.702 Percent Identity: 30.496
 alignment_block:
 US-08-957-709-70 x Q9PFB6
 Align seg 1/1 to: Q9PFB6 from: 1 to: 191
 76 GGTATGACCTCAGCTGAGTGGGAGAGAG..... 102
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 45 GlyTyraSpValArgCysSerArgGluPheLysilePheThrAsnIleAs 61
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 103GCTTTGTTAAGGGGAAATTAATCGACG 130
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 61 nSerThrIleValAspProLysGlnPheAspAsnGlySerPheIleAspv 78
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 131 TGGAAAGGAGGAGGAGGAGTCTTATTCCTCCAGGGAATACGCTTAATC 180
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 78 alGluSerAsp...ValcysIleIleProProAsnSerPheAlaLeuAla 93
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 181 CTAACCTCGAGAGGATAAAGTTGCCGACGATGTTATGGGGATATGAA 230
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 94 ArgThrIleGluTyrlPheArgIleProArgAsnValIleCysLe 110
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 231 GATAAGGAGCAGTTTACGAGAGAGAGGTTATTGTTCTTTTCCTTTGGG 280
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 110 uGlyLysSerThrTyrlAlaArgCysGlyIleIleValAsnValThrProL 127
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 281 TTGACCCAGGATGGGATGAGAACTTACACTAATGCTCTACATGCCTCA 330
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 127 euGluProGluTrpGluGlyHisValThrLeuGluPheSerAsnThrThr 143
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 331 AATGAACCTGTCGAAATTAAGATATGAGAGAGATTTGTGAGATCGCAT 380
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 144 ProLeuProAlaArgIleTyrlAlaAsnGluGlyValAlaGlnMetLeuPh 160
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 381 TATAAGGCTA.....GAGGGTCGGCAAGAACCCCTTACAGA..... 417
 :|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 160 eLeuGlnAlaAspProAspValCysGlnThrSerTyrlArgAspArg 177
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 418 ..GGAACTATCAGGGGAGGACACA 438
 |||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
 177 snGlyLysTyrlGlnGlyGlnThr 184
 seq_name: sp_archaea:Q9YG32
 seq_documentation_block:
 ID Q9YG32 PRELIMINARY; PRT; 163 AA.
 AC Q9YG32;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE 163AA LONG HYPOTHETICAL DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.
 GN APE0069.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix K1";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000058; BAA78978.1;
 DR InterPro; IPR003232; dCTP_deaminase.

```

DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Complete proteome.
SQ SEQUENCE 163 AA; 17384 MW; 4AA22FB0D8802F49 CRC64;

alignment_scores:
  Quality: 159.00      Length: 147
  Ratio: 1.916        Gaps: 5
  Percent Similarity: 56.463      Percent Identity: 32.653

alignment_block:
US-08-957-709-70 x Q9JG32 ..
Align seg 1/1 to: Q9JG32 from: 1 to: 163

37 CTTATAGACCATTTCTGAGATCGCTCAACGACGAGCTTATGACCT 86
   ::::::::::: ::::::::::: ::::::::::: :::::::::::
17 ValVallyGlyHisSerAsnGlyAlaIleGlnProAlaGlyValAspLe 33
87 CAGAGTGGGCAGA..... 99
33 userValGlyGluLeuGluSerLeuAlaAspAlaGlyPheLeuGlyGluG 50
100 .....GAGGCTTTGTTAAGGGGAAATTAATCGACGTGGAAAGGA 144
50 luAspLysIleMetProLysGlyAspArgIleGlnCysGlu...TyrGly 65
145 AAAGTCGTATTCTCTCAAGGGAATACGCCCTTAATCCTAACCCCTCGAGAG 194
66 ValCysGluLeuGluProGlyAlaTyrArgLeuArgPheAsn...Gluva 81
195 GATAAAGTTGCCCGACGATGTATGGGGATATGAAGTAAAGGACGATT 244
81 lvalSerIleProGlyHisValGlyPheCysPheProArgSerL 98
245 TAGCAAGAGAGGG...GTTATTGGTTCTTTGCTGGTTGACCCAGGA 291
98 euLeuArgMetGlyCystyrLeuGlyCysAlaValTrp...AspProGly 113
292 TGGGATCGAACTTAACACTAATGCTCTACAAATGCCCTCAATGAACCTGT 341
114 TyrThrGlyArgGlyGlnAlaMetLeuValAlaAsnProHisGlyLe 130
342 CGAATTAAGATATGAGAGAGATTGTGCAGATCGCATTTATAGGCTAG 391
130 uArgLeuGluMetGlySerArgIleAlaGlnLeuValValAlaArgValG 147
392 AGGTCCGGCAAGAAACCTTACAGAGAACTATCAGGGG 432
147 LuGlyProLeuThrSerLeuTyrLysGlyAspTyrGingly 160

seq_name: sp_bacteria:Q9JRE8

seq_documentation_block:
ID Q9JRE8 PRELIMINARY; PRT; 188 AA.
AC Q9JRE8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN NMA1060 (DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE,
DE PUTATIVE).
GN NMA1060 OR NMB0849.
OS Neisseria meningitidis (serogroup A), and
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699, 491,
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

```

```

RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Bartell B.G.;
RT *Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491";
RL Nature 404:502-506(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Feden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT *Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815(2000).
DR EMBL; AL162755; CAB84324.1; -.
DR EMBL; AE002438; AAF41260.1; -.
DR TIGR; NMB0849; -.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 188 AA; 21295 MW; 2CA4459F4E78FDF6 CRC64;

```

```

alignment_scores:
  Quality: 155.00      Length: 170
  Ratio: 1.566        Gaps: 4
  Percent Similarity: 58.235      Percent Identity: 24.118

alignment_block:
US-08-957-709-70 x Q9JRE8 ..
Align seg 1/1 to: Q9JRE8 from: 1 to: 188

37 CTTATAGACCATTTCTGAGATCGCTCAACGACGA..... 75
   ::::::::::: ::::::::::: :::::::::::
18 MetIleAspPropheGluProAsnGlnIleLysGluAlaAspGlyLysAr 34
76 .....GGTTATGACCTCAGAGTGGGCA 97
34 gIleIleSerTyrGlyThrSerTyrGlyTyrAspIleArgCysAlaA 51
98 GAGAG.....GCTTTTCTTAAGGGGAAATTAATCGACGCGAA 135
51 snGluPheLysIlePheThrAsnIleAsnSerThrIleValAspProLys 67
136 .....AAGGAAGGAAAGTCGTAT 155
68 AsnPheAspProLysAsnPheValThrValGluAspAspCysCysIleI 84
156 TCCCTCAAGGGAATACGCTTAATCCTAACCTCGAGAGGATAAGTTGC 205
84 eProProAsnSerPheAlaLeuAlaArgThrValGluTyrPheArgIleP 101
206 CCGACGATGTTATGGGGATATGAAGTAAAGGACGACTTTAGCAAGAA 255
101 roArgAsnValLeuThrValCysLeuGlyLysSerThrTyrAlaArgCys 117
256 GGGGTTATGCTTCTTTTGGTTGGTTGACCCAGGATGGATGCAACTT 305
118 GlyIleIleValAsnValThrProPheGluProGluTyrGlyTyrVa 134
306 AACACTAATGCTCTACAAATGCCTCAATGAACCTGTCGAATTAAGATG 355

```


This Page Blank (uspto)

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525


```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL445064; CAC11737.1; -.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
DR Hydrolase: Complete proteome.
KW SEQUENCE 183 AA; 20245 MW; 41D1992A99CD6682 CRC64;
SQ
alignment_scores:
  Quality: 257.50      Length: 155
  Ratio: 2.500         Gaps: 5
Percent Similarity: 66.452 Percent Identity: 41.290
alignment_block:
US-08-957-709-70 x DCD_THDAC ..
Align seg 1/1 to: DCD_THDAC from: 1 to: 183
1 ATGCTACTTCAGACTGGAATCAGAAAGAAATA..... 36
18 MetIleLeuAsnAspSerThrIleMetArgMetValSerAspGlyLeuLe 34
37 .CTTATAGAGCATTCTTCTGAAGAAATCCCAACCCAGCAGGTATGACC 85
34 uIleSerGluAsnPheAspArgGlyCysLeuThrProAsnGlyTyrAspL 51
86 TCAGAGTGGCGGAGAGAGCGCTTTTGTAGGGGAAATTAATCAGCGTGAA 135
51 euArgVal.....AspAlaIleAspValGlu 59
136 .....AAGGAGAAAGTCGTTATTCTCCCAAGGNAATACGCCTTAAT 179
60 GlyArgGlnTyrSerGluPheGluIleGlyLysAsnValHisPheLeuVa 76
180 CCTAACCTCGAGAGGATAAAGTTGCGCGACGATGTTATGGGGATATGA 229
76 IserThrIleGluLeuLysIleProAspAspValGlyMetIleT 93
230 AGATAAGAGCAGTTTACGAAGAAGGGTTATGTTCTTTGCTTGG 279
93 rpThrArgSerSerPheAlaArgLysGlyIlePheGlySerPheGlyAla 109
280 GTTGACCCAGGATGGTGAACCTTAACACTAATGCTCTACATGCTC 329
110 IleAspAlaGlyTyrHisGlyAsnLeuThrLeuSerPhePheAsnAlaGl 126
330 AAATGAACCTCTGAATAAGATATGGAGAGAGATTGCTGAGATCGCAT 379
126 ySer...AlaValAsnLeuArgArgGlyGluArgIleAlaGlnIleValP 142
380 TTATAAGCTAGAGGGTCCGACGAACCCCTTACAGA.....GGA 420
142 heValLysMetIleGlySerAlaGluLysProTyrHisIleArgSerGly 158
421 AACTATCAGGGGAGC 435
159 AsnTyrGlnAsnSer 163
seq_name: SwissProt_39:DCD_AQUAE
seq_documentation_block:
ID DCD_AQUAE STANDARD; PRT; 180 AA.
AC O67539;
```

```
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR AQ_1607.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.;
RL Nature 392:353-358(1998).
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000747; AAC07499.1; -.
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 180 AA; 20544 MW; B2710421A2FA48D6 CRC64;
```

```
alignment_scores:
  Quality: 243.00      Length: 162
  Ratio: 2.229         Gaps: 5
Percent Similarity: 67.284 Percent Identity: 37.654
alignment_block:
US-08-957-709-70 x DCD_AQUAE ..
Align seg 1/1 to: DCD_AQUAE from: 1 to: 180
1 ATGCTACTTCAGACTGGAATCAGAAATA.....GAAAT 35
1 MetIleLeuSerAspArgSerIleArgGluLeuLeuGlyGlyGluLe 17
36 ACTTATAGAGCATTCTTCTGAAGAAATCGCTCAACCCAGCAGGTATGACC 85
17 uLysValGluProTyrGluProSerHisValGlnCysSerSerLeuAspL 34
86 TCAGAGTGGCGGAGAG...GCTTTTGTTAAGGGGAAA...TTAATCGAC 129
34 euArgLeuGlyAsnGlnIleAlaLeuTyrGluGlyGluGlyValIleAsp 50
130 GTGAAAAGGAAGGAAAGTCGTT..... 153
51 ValLysLysGlyThrLysGlyValArgIleLeuGluGluTyrPh 67
154 ....ATTCTCCAAAGGAATACGCCTTAATCCTAACCCTCGAGAGATAA 199
67 eAspIleMetProLysGlnPheLeuLeuAlaThrThrLeuGluTyrIle 84
200 AGTTGCCCGGAGGTGTTATGGGGGATATGAAGATAAGCAGCAGTTTACCA 249
84 erLeuProProTyrValThrAlaPheValGluGlyArgSerSerLeuGly 100
```

250 AGAAGAGGGTT...ATTGGTCTTTCTTGGGTGACCCAGGATGGA 296
 101 ArgLeuGlyLeuPheIleGluAsnAlaGlyTrpValAspAlaGlyPheG 117
 297 TGGAACTAACACTAATGCTTACAAATGCCTCAATGAACGTGCGAAT 346
 117 uGlyGlnIleThrLeuGluLeuPheAsnAlaAsnAspArgProIleArgL 134
 347 TAAGATATGGAGAGAGATTGTGTCAGATCGCATTTATAAGGCTAGAGG 396
 134 euTyArgGlyMetArgIleCysGlnLeuValPheAlaArgLeuAspArg 150
 397 CCGGCAAGAACCCCTTACAGAGAACTATCAGGGG 432
 151 ProProGluArgValTySerGlyLysTyLysGly 162

seq_name: SwissProt_39:DCD_ACTIAM

seq_documentation_block:
 ID DCD_ACTIAM STANDARD; PRT; 173 AA.
 AC Q02103;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
 DE DEAMINASE).
 GN DCD.
 OS Acidianus ambivalens (Desulfurolobus ambivalens).
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Acidianus.
 ON NCBI_TaxID=2283;
 RN [1]
 RC STRAIN=Lei 10 / DSM 3772;
 RX MEDLINE=93065206; PubMed=1437556;
 RA Kletzin A.;
 RT "Molecular characterisation of a DNA ligase gene of the extremely
 RT thermophilic archaeon Desulfurolobus ambivalens shows close
 RT phylogenetic relationship to eukaryotic ligases.";
 RL Nucleic Acids Res. 20:5389-5396(1992).
 RP [2]
 RN SIMILARITY.
 RX MEDLINE=95206934; PubMed=7899076;
 RA Ouzounis C., Kyriades N., Sander C.;
 RT "Novel protein families in archaean genomes.";
 RL Nucleic Acids Res. 23:565-570(1995).
 CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
 CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X63438; CAA45033.1;
 DR PIR: S26382; S26382;
 DR InterPro: IPR003232; dCTP_deaminase.
 DR InterPro: IPR001428; dUTPase.
 DR Pfam: PF00652; dUTPase; 1.
 DR ProDom: PD004900; dCTP_deaminase; 1.
 KW Hydrolyase.
 SQ SEQUENCE. 173 AA; 19858 MW; B4D922503CD4B25A CRC64;

alignment_scores:
 Quality: 242.00 Length: 156
 Ratio: 2.283 Gaps: 3
 Percent Similarity: 67.949 Percent Identity: 35.897

alignment_block:
 US-08-957-709-70 x DCD_ACTIAM ..

Align seg 1/1 to: DCD_ACTIAM from: 1 to: 173

34 ATACTTATAGAGCCATTTCTGAAGATCGCTCAACAGCAGGTTATGA 83
 17 IleValIleSerProLeuThrGluAspThrIleArgGluAsnGlyValas 33
 84 CCTCAGAGTGGGC.....AGAGAGGCTT 106
 33 PleuArgValGlyGlyGluIleAlaArgPheLysLysThrAspGluIleT 50
 107 TGTGTAAGGGGAAA.....TTAATCGAGCTGGAAGAAAGGA 144
 50 YrGluAspGlyLysAspProArgSerPheTyrgluIleGluLysGlyasp 66
 145 AAGTCGTATTCTCCCAAGGAATACGCTTAATCTTAACCTCGAGAG 194
 67 GlupheIleIleTyProAsnGluHisValLeuLeuValThrGluGluTy 83
 195 GATAAAGTTGCCGACGATGTATGGGGATATGAAGATAAGGACGATT 244
 83 rValLysLeuProAsnAspValMetAlaPheValAsnLeuArgSerP 100
 245 TAGCAAGAAGAGGGTTATTCGTTCTTTGCTGGTTGACCCAGGATGG 294
 100 heAlaArgLeuGlyLeuPheValProThrIleValAspAlaGlyPhe 116
 295 GATGAAACTTAACACTAATCTCTACAAATGCCTCAATGAACCTGCGA 344
 117 GluGlyGlnLeuThrIleGluValLeu...GlySerAlaPheProVally 132
 345 ATTAAGATATGGAGAGAGATTGTGCAGATCGCATTTATTAAGGCTAGAG 394
 132 sileLysArgGlyThrArgPheLeuHisLeuIlePheAlaArgThrLeu 149
 395 GTCCGGCAAGAAACCTTACAGAGAACTATCAGGGAGCAGCAAGGTTA 444
 149 hrProValGluAsnProTyHisGlyLysTyrglnGlyGlnGlnGlyVal 165
 445 CGGTTTCAAAGAGAAAG 462
 166 ThrLeuProLysPheLys 171

seq_name: SwissProt_39:DCD_AERPE

seq_documentation_block:
 ID DCD_AERPE STANDARD; PRT; 181 AA.
 AC Q9YFAB;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
 DE DEAMINASE).
 GN DCD OR APE0133.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kwarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
 CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000059; BAA79288.1;
DR InterPro; IPR003232; dCTP_deaminase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
DR Hydrolase; Complete proteome.
SQ SEQUENCE 181 AA; 19894 MW; D8B6CBDC1722EFE9 CRC64;

alignment_scores:
  Quality: 235.00      Length: 174
  Ratio: 2.136        Gaps: 5
  Percent Similarity: 63.218      Percent Identity: 34.483

alignment_block:
US-08-957-709-70 x DCD_AERPE ..
Align seg 1/1 to: DCD_AERPE from: 1 to: 181

1 ATGCTACTTCAGACTGGAATCAGAAA.....GAAAT 35
4 LeuLeuSerAspArgAspIleArgAlaLeuLeuAlaIleGlyAspLe 20
36 ACTTATAGAGCCATTCTTGAAGAACTGCTCCCAACAGAGGTTATGACC 85
20 uValValGluProLeuSerGlyAspThrValArgGluAsnGlyLeuAspL 37
86 TCAGAGGGGCGAGAGCGCTTTTCTG..... 110
37 euArgLeuGly..ArgGlyPheCysArgPheLysArgSerAspArgValL 53
111 .....TAAGGGGAAATTAATCGACGTGGAAA 137
53 euAspProArgAlaProGlySerProGlyGluPheTyrGluCysGlyGlu 69
138 GGAAGGAAAGTCGTTATCTCTCAAGGGAATCGCTTAACTCCTAATCCTAACCC 187
70 Gly ASPGluIleValGlyProGlyGluHisMetLeuLeuHisThrG 86
188 TCAGAGAGGATAAAGTCCCGACGATGTTATGGGGGATATGAAGTAAGG 237
86 InGluTyrIleArgLeuProGlyTyrValAlaGlyLeuValAsnLeuArg 102
238 AGCAGTTTACGAGAGAGGGTTATGTTCTTTCTGTTGGGTTGACCC 287
103 SerThrTrpAlaArgThrGlyIleTyrIleProAlaThrValValAspAl 119
288 AGGATGGATGGAACCTAACACTAATGCTCAATGCCTCAATGAAC 337
119 aGlyPheGluGlyGlnLeuIleGluValVal...GlySerGlyPheP 135
338 CTGTGCAATTAAGATATGAGAGAGATTGTGTCAGATCGCATTTATAAGG 387
135 roValLysLeuTyrProGlyAspArgPheLeuHisLeuValLeuValLys 151
388 CTAGAGGGTCGGCAGAAACCCCTTACAGAGGAACCTATCAGGGGAGCAC 437
152 LeuGlnSerProAlaMetAsnProTyrArgGlyArgTyrGlnGlyGlnAr 168
438 AAGCTTAGCGCTTTTCAAG 456
168 gGlyValArgLeuProLys 174
seq_name: SwissProt_39:DCD_STRCO
```

```
seq_documentation_block:
ID DCD_STRCO STANDARD; PRT; 191 AA.
AC Q9X8W0;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR SCH35.46
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,
RA Rajandream M.A.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O = dUTP + NH(3).
CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL078610; CAB44381.1;
DR InterPro; IPR003232; dCTP_deaminase.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
DR Hydrolase.
SQ SEQUENCE 191 AA; 21496 MW; 6352A496990F910C CRC64;
```

```
alignment_scores:
  Quality: 216.00      Length: 169
  Ratio: 2.077        Gaps: 3
  Percent Similarity: 61.538      Percent Identity: 31.361

alignment_block:
US-08-957-709-70 x DCD_STRCO ..
Align seg 1/1 to: DCD_STRCO from: 1 to: 191

1 ATGCTACTTCAGACTGGAATCAGAAAAGAAATA..... 36
1 MetLeuLeuSerAspLysAspIleArgAlaGluIleAspAsnGlyArgVa 17
37 .CTTATAGAGCCATTCTCTGAAGAACTGCTCCCAACAGAGGTTATGACC 85
17 lArgIleAspProPheAspSerMetValGlnProSerSerIleAspV 34
86 TCAGAGTGGCGCAGA..... 99
34 aArgLeuAspArgTyrPheArgValPheGluAsnHisArgTyrProHis 50
100 .....GAGGCTTTTGTAAAGGGGAAATTAATCGACGTGGA 134
51 IleAspProSerValGluGlnValAspLeuThrArgLeuValGluProGl 67
135 AAAGAGAGGAAAGTCGTTATCTCTCAAGGGAATACGCTTAACTCTAA 184
67 uGlyAspGluProPheIleLeuHisProGlyGluPheValLeuAlaSerT 84
185 CCCTCGAGAGGATAAAGTCCCGACGATGTTATGGGGGATATGAAGATA 234
84 hrTyrGluValValSerLeuProAspAspLeuAlaSerArgLeuGluGly 100
```



```

RX MEDLINE-98295987; PubMed-9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z96800; CAB09605.1; -.
DR EMBL; AE006940; AAK44559.1; -.
DR HSSP; P16088; 1DUT.
DR TIGR; MT0336; -.
DR Tuberculist; Rv0321; -.
DR InterPro; IPR003232; dCTP_deaminase.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD004900; dCTP_deaminase; 1.
DR Hydrolase; Complete proteome.
SQ SEQUENCE 190 AA; 20869 MW; F409329810B64781 CRC64;

alignment_scores:
  Quality: 198.00      Length: 169
  Ratio: 1.941        Gaps: 3
Percent Similarity: 60.355 Percent Identity: 30.178

alignment_block:
US-08-957-709-70 x DCD_MYCTU ..
Align seg 1/1 to: DCD_MYCTU from: 1 to: 190

1 ATGCTACTTCAGACTGGAATATCGAAAGAAATACCTT..... 39
|||||  |||  |||  |||  |||  |||  |||  |||  |||
1 MetLeuLeuSerArgAspLeuArgAlaGluLeuSerSerGlyArgLe 17

40 ....ATAGGACCATTTCTGAGATCGCTCCACACGACGATTATGACC 85
|||||  |||  |||  |||  |||  |||  |||  |||  |||
17 uGlyIleAspPropheAspThrLeuValGlnProSerIleAspV 34

86 TCAGAGTG..... 93
|||||  |||  |||  |||  |||  |||  |||  |||  |||
34 alArgLeuAspCysLeuPheArgValPheAsnThrArgTyrThrHis 50

94 .....GGCAGAGAGCGTTTGTAAAGGGGAAATTAATCGAGCGTGA 134
|||||  |||  |||  |||  |||  |||  |||  |||  |||
51 IleAspProAlaLysGlnAspGluLeuThrSerLeuValGlnProVa 67

```

```

135 AAAGGAAGAAAGTCGTATTTCCTCAAGGAATACCCCTTAATCCTAA 184
|||||  |||  |||  |||  |||  |||  |||  |||  |||
67 lAspGlyGluProPheValLeuHisProGlyGluPheValLeuGlySert 84

185 CCCTCGAGAGATAAAGTTGTCGCCGACGATGTATGGGGATATGAGATA 234
|||||  |||  |||  |||  |||  |||  |||  |||  |||
84 hrLeuGluLeuPheThrLeuProAspAsnLeuAlaGlyArgLeuGluGly 100

235 AGGACGACCTTAGCAAGAGAAAGGGGTATT.....GGTCTTTTGGCTTG 278
|||||  |||  |||  |||  |||  |||  |||  |||  |||
101 LysSerSerLeuGlyArgLeuLeuThrHisSerThrAlaGlyPh 117

279 GGTGACCCAGATGGGATGAAACTTAACACTAATGCTACATATGCT 328
|||||  |||  |||  |||  |||  |||  |||  |||  |||
117 eIleAspProGlyPheSerGlyHisIleThrLeuGluLeuSerAsnVala 134

329 CAATGACACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGATCGCA 378
|||||  |||  |||  |||  |||  |||  |||  |||  |||
134 laAsnLeuProIleThrLeuTrpProGlyMetLysIleGlyGlnLeuCys 150

379 TTTATAAGGCTAGAGGTCGCCGCAAGAAACCTTACAGAGGAAACTATCA 428
|||||  |||  |||  |||  |||  |||  |||  |||  |||
151 MetLeuArgLeuThrSerProSerGluHisProGlySerSerArgAl 167

429 GGGGAGC 435
|||||
167 aGlySer 169

seq_name: SwissProt_39:DCD_HELPY
seq_documentation_block:
ID DCD_HELPY STANDARD; PRT; 188 AA.
AC 025136;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR HP0372.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE-97394467; PubMed-9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson J., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000554; AAD07441.1; ALT_INIT.
DR TIGR; HP0372; -.
DR InterPro; IPR003232; dCTP_deaminase.

```

```
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase. 1.
DR ProDom: PD004900; dCTP deaminase; 1.
RW Hydrolase; complete proteome.
SQ SEQUENCE 188 AA; 20904 MW; F45ABE3C3F040876 CRC64;

alignment_scores:
  Quality: 181.00      Length: 168
  Ratio: 1.782        Gaps: 4
  Percent Similarity: 60.119      Percent Identity: 27.976

alignment_block:
  US-08-957-709-70 x DCD_HELPY
  Align seg 1/1 to: DCD_HELPY from: 1 to: 188

37 CTTATAGAGCCATTTCTGAAGAA..... 60
18 MetIleSerProPheCysGluLysValGlyLysAsnValIleSerTy 34
61 .TCGCTCCAAACCAGCAGGTTATGACCTCAGAGTGGCGCAGAG..... 102
34 rGlyLeuSerSerTyGlyTyAspIleArgValGlySerGluPheMetL 51
103 .....GCTTTTGTAAAGGGGAAA..... 120
51 eUpheAspAsnLysAsnAlaLeuIleAspProLysAsnPheAspProAsn 67
121 .....TTAATCGACGTGAAAGGAAAGGAAAGTCTGTTATTCCTCC 161
68 AsnAlaThrLysIleAspAlaSerLysGluGlyTyPheIleLeuProAl 84
162 AAGGAATACGCTTAATCCTAACCTCGAGAGGATAAGTTGCCCGACG 211
84 aAsnAlaPheAlaLeuAlaHisThrIleGluTyPheLysMetProLysA 101
212 ATGTTATGGGGATATGAAGTAAGACCTGTCGAATTAAGATATGAGAGA 361
134 eLulIleSerAsnThrThrAsnLeuProAlaLysValTyAlaAsnGluG 151
362 GATTGTGCAGATCCGATTTAAGGCTAGAGGTCGGCGCAAGAACCT 411
151 lYleAlaGlnValPheLeuGlnGlyAspGluMetCysGluGlnSer 167
412 TACAGA.....GGAACATATCAGGGGAGCCAGGTTAGCGTTTC 452
168 TyrLysAspArgGlyGlyLysTyrglnGlyGlnValGlyIleThrLeuPr 184
453 AAAG 456
184 olys 185

seq_name: SwissProt_39:DCD_HELPY
seq_documentation_block:
  AC Q9ZKD0;
  DT 20-AUG-2001 (Rel. 40, Created)
  DT 20-AUG-2001 (Rel. 40, Last sequence update)
  DT 20-AUG-2001 (Rel. 40, Last annotation update)
  DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
  DE DEAMINASE).
  GN DCD OR JHP1009.
  OS Helicobacter pylori J99 (Campylobacter pylori J99).
```


[illegible]

24 sLysGlnLeuIleAsnGlyIleThrValAspIleHisLeuGlyAsnLysp 41
 102
 41 heArgPhePheTyAspHisThrThrSerCysIleAspLeuSerGlySer 57
 103GCTTTTGTAAAGGGGAAATTAATCAGCGTGGAA..... 135
 58 LysGluLysIleAlaLeuAspLeuAsnLysIleValSerCysGluThrIle 74
 136AAGGAAGAAAGTCGTATTATCCCAAGGAATACGCCCTAA 178
 74 ePheSerLysLysGluProPhePheLeuLysProGlyAlaLeuAlaLeup 91
 179 TCCTAACCCCTGAGAGGATTAAGTTCGCCGACGATGTTATGGGGGATATG 228
 91 heSerThrLeuGluAsnIleThrLeuProAsnAsnLeuValGlyTrpLeu 107
 229 AAGATAAGGAGCAGTTAGCAAGAGAAGGGGTATTGCTTTTGTCT... 276
 108 AspGlyArgSerSerLeuAlaArgLeuGlyLeuMetValHisValThrSe 124
 277TGGGTGACCCAGATGGGATGAACCTTAACACTAATGCTCTACA 322
 124 rHisArgIleAspProGlyTrpHisGlyAsnIleValLeuGluPhePheA 141
 323 ATGCTCAATGAACCTGTCCGAATTAAGATATGAGAGAGATTTGTGCAG 372
 141 snAlaGlyLysLeuThrValLeuValLeuThrProGlyIleLysIleAlaAla 157
 373 ATCCGATTTTAAAGCGGTAGAGGGTCCGCGAAGAAACCTTCACAGAGAAA 422
 158 LeuSerPheGluLeuLeuSerLysProValLeuArgProTyAsnSerAr 174
 423 CTATCAGGGAGCACAGGTAGCGTTTCAACAGAGAAGAAA 465
 174 gAsnGlu.....SerLysTyLysArg 181

seq_name: SwissProt_39:DCD_METJA

seq_documentation_block:

ID DCD_METJA STANDARD; PRT; 204 AA.
 AC 057872;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
 DE DEAMINASE).
 GN DCD OR MJ0430.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
 CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U67494; AAB98415.1; -.
 DR TIGR: MJ0430; -.
 DR InterPro: IPR003232; dCTP_deaminase.
 DR InterPro: IPR001428; dUTPase.
 DR Pfam: PF00692; dUTPase; 1.
 DR ProDom: PD004900; dCTP_deaminase; 1.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 204 AA; 23432 MW; 1218368057723371 CRC64;

alignment_scores:

Quality: 154.00 Length: 183
 Ratio: 1.510 Gaps: 6
 Percent Similarity: 55.738 Percent Identity: 25.683

alignment_block:

US-08-957-709-70 x DCD_METJA ..

Align seg 1/1 to: DCD_METJA from: 1 to: 204

13 CACTGGAAAAATCAGAAAGAAATCTTATAGAGCCATTTCTGAAGAATC 62
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 10 AspTyValThrSerLysArgIleIleIleLysProPheAsnLysAsp 26
 63 GCTCAACACGAGGTATGACCTCAGAGTGGCAGAGAGCGCTTTTGT 111
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 26 eValGlyProCysSerTyAspValThrLeuGlyAspGluPheIleIle 43
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 112 ..AAGGGGAAATTAATCGACCTGGAAGAA..... 141
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 43 TyrAspGluValTyAspLeuSerLysGluLeuAsnTyLysArgIle 59
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 142GGAAGTCGTTATTCCTCAAGGAATACGCTTAATC... 180
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 60 LysIleLysAsnSerIleLeuValCysProLeuAsnTyAsnLeuThrG 76
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 180 180
 76 uGluLysIleAsnTyPheLysGluLysTyAsnValAspTyValValG 93
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 181CTAACCTCGAGAGGATAAAGTTCGCCGACGAT 213
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 93 LuGlyGlyValLeuGlyThrThrAsnGluTyIleGluLeuProAsnAsp 109
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 214 GTTATGGGGGATATCAAGATAGGAGCAGTTTACGAGAGAGGGGTTAT 263
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 110 IleSerAlaGlnTyArgSerSerLeuGlyArgValPheLeuTh 126
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 264 TGGT.....TCTTTTGGTTCACCCAGAGTGGATGGAACTTAA 307
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 126 rSerHisGlnThrAlaGlyTrpIleAspAlaGlyPheLysGlyLysIle 143
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 308 CACTAATGCTCTACAATGCCCTCAATGAACCTGCGAATTAAGATATGGA 357
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 143 hrLeuGluIleVal...AlaPheAspLysProValIleLeuTyLysAsn 158
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 358 GAGAGATTGTGCGAGTCCGATTTATAGGCTAGAGGTCCGGCAGAGAAA 407
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 159 GlnArgIleGlyGlnLeuPheSerLysLeuLeuSerProAlaAspVa 175
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 408 CCCCTACAGAGAACTATACAGGGGAGCACAGGTTACGCTTTTCAAAG 456
 :||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 175 lGlyTy.....SerGluArgLysThrSerLysTyAlaTyGlnLys 189

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 10:47:47 ; Search time 38.81 seconds
(without alignments)
2748.548 Million cell updates/sec

Title: US-08-957-709-70
Perfect score: 471
Sequence: 1 ATGCTACTTCCAGACTGGAA.....CAAGAGAAAGAACTCTAG 471

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCRUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	316	67.1	740	4	US-08-822-774-42	Sequence 42, Appl
C 2	96.4	20.5	129	4	US-08-822-774-40	Sequence 40, Appl
3	82	17.4	84	4	US-08-822-774-36	Sequence 36, Appl
C 4	54.2	11.5	7218	1	US-08-232-463-14	Sequence 14, Appl
C 5	31.8	6.8	491	4	US-08-896-164-11	Sequence 11, Appl
6	31.4	6.7	2646	1	US-08-365-189-7	Sequence 7, Appl
7	31.4	6.7	2982	1	US-08-365-189-2	Sequence 2, Appl
8	31.4	6.7	2988	1	US-08-365-189-1	Sequence 1, Appl
C 9	31.2	6.6	802	4	US-08-896-164-12	Sequence 12, Appl
10	31.2	6.6	960	2	US-08-800-264A-2	Sequence 2, Appl
11	31.2	6.6	960	2	US-09-018-628-2	Sequence 2, Appl
12	31.2	6.6	960	3	US-09-273-378-2	Sequence 2, Appl
13	31.2	6.6	960	3	US-09-209-605-2	Sequence 2, Appl
14	31.2	6.6	1311	2	US-08-800-264A-3	Sequence 3, Appl
15	31.2	6.6	1311	2	US-09-018-628-3	Sequence 3, Appl
16	31.2	6.6	1311	3	US-09-273-378-3	Sequence 3, Appl
17	31.2	6.6	1311	3	US-09-209-605-3	Sequence 3, Appl
18	31.2	6.6	1317	2	US-08-800-264A-1	Sequence 1, Appl
19	31.2	6.6	1317	2	US-09-018-628-1	Sequence 1, Appl
20	31.2	6.6	1317	3	US-09-273-378-1	Sequence 1, Appl
21	31.2	6.6	1317	3	US-09-209-605-1	Sequence 1, Appl
C 22	31.2	6.6	1581	5	PCT-US95-08493-14	Sequence 14, Appl
23	31.2	6.6	1629	1	US-08-519-103-11	Sequence 11, Appl
24	31.2	6.6	1629	4	US-09-018-635-11	Sequence 11, Appl
C 25	31.2	6.6	2580	5	PCT-US95-08493-18	Sequence 18, Appl
C 26	31.2	6.6	2604	5	PCT-US95-08493-20	Sequence 20, Appl
C 27	31.2	6.6	2686	3	US-09-358-384-1	Sequence 1, Appl

28	30.8	6.5	15328	2	US-08-888-497-33	Sequence 33, Appl
29	30.8	6.5	15328	5	PCT-US94-07926-33	Sequence 33, Appl
30	29.8	6.3	831	2	US-08-403-853-17	Sequence 17, Appl
31	29.8	6.3	9472	1	US-08-325-547-9	Sequence 9, Appl
32	29.6	6.3	381	1	US-08-137-117D-36	Sequence 36, Appl
33	29.6	6.3	381	1	US-08-436-717-36	Sequence 36, Appl
34	29.6	6.3	2363	2	US-08-671-978A-11	Sequence 11, Appl
35	29	6.2	2022	1	US-08-803-973-6	Sequence 6, Appl
36	29	6.2	2022	1	US-08-803-972-6	Sequence 6, Appl
37	29	6.2	2124	1	US-08-803-973-11	Sequence 11, Appl
38	29	6.2	2124	1	US-08-803-972-11	Sequence 11, Appl
39	29	6.2	4371	1	US-08-803-973-1	Sequence 1, Appl
40	29	6.2	4371	1	US-08-803-972-1	Sequence 1, Appl
41	28.4	6.0	2421	1	US-08-063-552-1	Sequence 1, Appl
42	28.4	6.0	2421	5	PCT-US93-05704-1	Sequence 1, Appl
43	28.4	6.0	2656	2	US-08-685-625A-5	Sequence 5, Appl
44	28.4	6.0	9495	1	US-08-271-829-1	Sequence 1, Appl
45	28.4	6.0	9495	5	PCT-US93-01544-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-822-774-42
; Sequence 42, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/822,774
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-822-774-42

Query Match 67.1%; Score 316; DB 4; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.3e-96;
Matches 316; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 156 TCCTCCAGGGAATACGCTTAACCTTAACCCCTCCAGAGATAAAGTTGCCCGACGATGT 215

1 TCCCTCCAGGATACGCTTAATCTAACCTCAGAGGATAAAGTTGCCGACGATGT 60
216 TATGGGGATATGAGATTAAGGACGCTTACGAGAGGAGGCTTATGCTCTTTGTC 275
61 TATGGGGATATGAGATTAAGGACGCTTACGAGAGGAGGCTTATGCTCTTTGTC 120
276 TTGGGTTGACCCAGGATGGGATGAACTTAACACTATGCTACAAATGCTCAATGA 335
121 TTGGGTTGACCCAGGATGGGATGAACTTAACACTATGCTACAAATGCTCAATGA 180
336 ACCTGTGCAATTAAGATGATGAGAGAGATTTGTGACAGATCCGATTTATPAAGGCTAGAGG 395
181 ACCTGTGCAATTAAGATGATGAGAGAGATTTGTGACAGATCCGATTTATPAAGGCTAGAGG 240
396 TCCGGCAAGAACCTTACAGAGGAACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAA 455
241 TCCGGCAAGAACCTTACAGAGGAACTATCAGGGGAGCACAAGGTTAGCGTTTTCAAA 300
456 GAGAAAGAACTCTAG 471
301 GAGAAAGAACTCTAG 316

RESULT 2
US-08-822-774-40/c
; Sequence 40, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-822-774-40

Query Match 20.5%; Score 96.4; DB 4; Length 129;
Best Local Similarity 99.0%; Pred. No. 2.9e-23;
Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGCTACTTCAGACTCGAAATCAGAAAGAAATATATAGAGCCATTTCTCGAAGAA 60
98 ATGCTACTTCAGACTCGAAATCAGAAAGAAATATATAGAGCCATTTCTCGAAGAA 39
61 TCGCTCCAACACGAGGTTATGACCTCAGAGTGGGCAG 98
38 TGGCTCCAACACGAGGTTATGACCTCAGAGTGGGCAG 1

RESULT 3
US-08-822-774-36
; Sequence 36, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; ANTI-SENSE: NO
US-08-822-774-36

Query Match 17.4%; Score 82; DB 4; Length 84;
Best Local Similarity 97.6%; Pred. No. 1.6e-18;
Matches 82; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 87 CAGAGTGGGCGAGAGGCTTTTGTAAAGGGGAAATTAATCAGCTGGGAAAGGAGGAAA 146
DB 1 CAGAGTGGGCGAGAGGCTTTTGTAAAGGGGAAATTAATCAGCTGGGAAAGGAGGAAA 60
QY 147 AGTCGTTATTCCTCCAAGGGAATA 170
DB 61 AGTCGTTATTCCTCCAAGGGAATA 84

RESULT 4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.


```

; TITLE OF INVENTION: Cloned Pullulanase
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/132,648
; FILING DATE: October 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ryser, David G.
; REGISTRATION NUMBER: 36,407
; REFERENCE/DOCKET NUMBER: 66-005-9367-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 271-5717
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-365-189-7

Query Match 6.7%; Score 31.4; DB 1; Length 2646;
Best Local Similarity 59.6%; Pred. No. 0.91; Mismatches 0; Indels 0; Gaps 0;
Matches 53; Conservative 0;

Qy 211 GATGTTATGGGGGATATGAAGATAGGAGCAGTTAGCAAGAGAGAGGGTTATTGGTTCT 270
Db 1408 GATCTTATGGGGCATATCATGAAAATACCATGATAGAGCAAAATCTGCTATTTCGAAGC 1467
Qy 271 TTTCGTTGGTTGACCCAGGATGGATGG 299
Db 1468 CTTACGAGGGATGTACATGGAGTGGATGG 1496

RESULT 7
US-08-365-189-2
; Sequence 2, Application US/08365189
; Patent No. 5514576
; GENERAL INFORMATION:
; APPLICANT: Bower, Patricia A.
; TITLE OF INVENTION: Cloned Pullulanase
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/132,648
; FILING DATE: October 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ryser, David G.
; REGISTRATION NUMBER: 36,407
; REFERENCE/DOCKET NUMBER: 66-005-9367-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 271-5717
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2988 base pairs
; TYPE: nucleic acid
; APPLICATION NUMBER: US/08/365,189

```

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/132,648
; FILING DATE: October 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ryser, David G.
; REGISTRATION NUMBER: 36,407
; REFERENCE/DOCKET NUMBER: 66-005-9367-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 271-5717
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2982 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Oligonucleotide
; US-08-365-189-2

Query Match 6.7%; Score 31.4; DB 1; Length 2982;
Best Local Similarity 59.6%; Pred. No. 0.97; Mismatches 0; Indels 0; Gaps 0;
Matches 53; Conservative 0;

Qy 211 GATGTTATGGGGGATATGAAGATAGGAGCAGTTAGCAAGAGAGAGGGTTATTGGTTCT 270
Db 1402 GATCTTATGGGGCATATCATGAAAATACCATGATAGAGCAAAATCTGCTATTTCGAAGC 1461
Qy 271 TTTCGTTGGTTGACCCAGGATGGATGG 299
Db 1462 CTTACGAGGGATGTACATGGAGTGGATGG 1490

RESULT 8
US-08-365-189-1
; Sequence 1, Application US/08365189
; Patent No. 5514576
; GENERAL INFORMATION:
; APPLICANT: Bower, Patricia A.
; TITLE OF INVENTION: Cloned Pullulanase
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/132,648
; FILING DATE: October 5, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ryser, David G.
; REGISTRATION NUMBER: 36,407
; REFERENCE/DOCKET NUMBER: 66-005-9367-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 271-5717
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2988 base pairs
; TYPE: nucleic acid

```


Db 618 AATAACTTCTCAAGATAAACCTAG 641

RESULT 11

US-09-018-628-2

; Sequence 2, Application US/09018628

; Patent No. 5917019

; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia

; APPLICANT: van Steensel, Bas

; APPLICANT: Bianchi, Alessandro

; TITLE OF INVENTION: AN ALTERED TELOMERE REPEAT BINDING

; TITLE OF INVENTION: FACTOR AND THERAPEUTIC USE THEREOF

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Jackson, Esq.

; STREET: 411 Hackensack Ave, Continental Plaza, 4th

; STREET: Floor

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/018,628

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-191 CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800

; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 960 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

US-09-018-628-2

Query Match 6.6%; Score 31.2; DB 2; Length 960;

Best Local Similarity 47.1%; Pred. No. 0.63; 108; Indels 0; Gaps 0;

Matches 96; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTTATAGAGCCATTTCTGAAGAA 60

Db 438 AAGCAATTTGCTTATGATATCTCTCAGAAAGATACATTTCTTTTCAACACTT 497

Qy 61 TCGTCCCAACCCAGCAGGTTATGACCTCAGAGTGGCGCAGAGAGGCTTTTGTAAAGGGAAA 120

Db 498 CAGCTACAACCCACATGATGGAGAAATTAAGAGTTATGTGAATTTATGTCTTAAGTGAAA 557

Qy 121 TTAATCGACGTGGAAAGAGGAAAGTCTGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180

Db 558 ATCAATCAACCTTTCTAATGAAGCGCGGCAAAAGTAGTAAAGCAAAAGGCAAGAAC 617

Qy 181 CTAACCCCTCGAGAGGATAAAGTTG 204

Db 618 AATAACTTCTCAAGATAAACCTAG 641

RESULT 12

US-09-273-378-2

; Sequence 2, Application US/09273378

; Patent No. 6020166

; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia

; APPLICANT: van Steensel, Bas

; APPLICANT: Bianchi, Alessandro

; TITLE OF INVENTION: AN ALTERED TELOMERE REPEAT BINDING

; TITLE OF INVENTION: FACTOR AND THERAPEUTIC USE THEREOF

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Jackson, Esq.

; STREET: 411 Hackensack Ave, Continental Plaza, 4th

; STREET: Floor

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/273,378

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/018,628

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-191 CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800

; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 960 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

US-09-273-378-2

Query Match 6.6%; Score 31.2; DB 3; Length 960;

Best Local Similarity 47.1%; Pred. No. 0.63; 108; Indels 0; Gaps 0;

Matches 96; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTTATAGAGCCATTTCTGAAGAA 60

Db 438 AAGCAATTTGCTTATGATATCTCTCAGAAAGATACATTTCTTTTCAACACTT 497

Qy 61 TCGTCCCAACCCAGCAGGTTATGACCTCAGAGTGGCGCAGAGAGGCTTTTGTAAAGGGAAA 120

Db 498 CAGCTACAACCCACATGATGGAGAAATTAAGAGTTATGTGAATTTATGTCTTAAGTGAAA 557

Qy 121 TTAATCGACGTGGAAAGAGGAAAGTCTGTTATTCCTCCAAAGGGAATACGCCCTTAATC 180

Db 558 ATCAATCAACCTTTCTAATGAAGCGCGGCAAAAGTAGTAAAGCAAAAGGCAAGAAC 617

Qy 181 CTAACCCCTCGAGAGGATAAAGTTG 204

Db 618 AATAACTTCTCAAGATAAACCTAG 641

RESULT 13

US-09-209-605-2

; Sequence 2, Application US/09209605

; Patent No. 6023709

; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia

APPLICANT: van Steensel, Bas
APPLICANT: Bianchi, Alessandro
TITLE OF INVENTION: AN ALTERED TELOMERE REPEAT BINDING
TITLE OF INVENTION: FACTOR AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/209,605
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/800,264
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-191
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
US-09-209-605-2

Query Match 6.6%; Score 31.2; DB 3; Length 960;
Best Local Similarity 47.1%; Pred. No. 0.63;
Matches 96; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Qy 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTTATAGAGCCATTTCTGAAGAA 60
Db 438 AAGCAAAATGCTTATGATAATCTCTCAGAAAGATACATTTTCATTCCTTTTCAACACTT 497
Qy 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGCAGAGAGGCTTTGTTAAGGGGAAA 120
Db 498 CAGCTACAAACCAATGATGGAGAAATTAAGAGTTATGTGCTAAGTGAAGAA 557
Qy 121 TTAATCGACGTGGAAAGAGAAAGTTCGTTATTCCTCCAAAGGGAATAGCCCTTAATC 180
Db 558 ATCATCAACCTTTCTAATGAAGGAGCGGCAAAAGTAGTAGAAGCAAGCAAGAAC 617
Qy 181 CTAACTCTCGAGAGATAAAGTTG 204
Db 618 AATACTTCTCAAGATAAACCTAG 641

RESULT 14
US-08-800-264A-3
Sequence 3, Application US/08800264A
Patent No. 5859183
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: van Steensel, Bas
APPLICANT: Bianchi, Alessandro
TITLE OF INVENTION: AN ALTERED TELOMERE REPEAT BINDING

TITLE OF INVENTION: FACTOR AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,264A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-191
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
US-08-800-264A-3

Query Match 6.6%; Score 31.2; DB 2; Length 1311;
Best Local Similarity 47.1%; Pred. No. 0.74;
Matches 96; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Qy 1 ATGCTACTTCCAGACTGGAAATCAGAAAGAAATACCTTATAGAGCCATTTCTGAAGAA 60
Db 633 AAGCAAAATGCTTATGATAATCTCTCAGAAAGATACATTTTCATTCCTTTTCAACACTT 692
Qy 61 TCGCTCCAAACAGCAGGTTATGACCTCAGAGTGGCAGAGAGGCTTTGTTAAGGGGAAA 120
Db 693 CAGCTACAAACCAATGATGGAGAAATTAAGAGTTATGTGCTAAGTGAAGAA 752
Qy 121 TTAATCGACGTGGAAAGAGAAAGTTCGTTATTCCTCCAAAGGGAATAGCCCTTAATC 180
Db 753 ATCATCAACCTTTCTAATGAAGGAGCGGCAAAAGTAGTAGAAGCAAGCAAGAAC 812
Qy 181 CTAACTCTCGAGAGATAAAGTTG 204
Db 813 AATACTTCTCAAGATAAACCTAG 836

RESULT 15
US-09-018-628-3
Sequence 3, Application US/09018628
Patent No. 5917019
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: van Steensel, Bas
APPLICANT: Bianchi, Alessandro
TITLE OF INVENTION: AN ALTERED TELOMERE REPEAT BINDING
TITLE OF INVENTION: FACTOR AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

;; CITY: Hackensack
;; STATE: New Jersey
;; COUNTRY: USA
;; ZIP: 07601
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patent in Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/018,628
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-191 CIP
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 201-487-5800
;; TELEFAX: 201-343-1684
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1311 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
US-09-018-628-3

Query Match 6.6%; Score 31.2; DB 2; Length 1311;
Best Local Similarity 47.1%; Pred. NO. 0.74;
Matches 96; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 1 ATGCTACTTCCAGACTGGAAAATCAGAAAAGAAATACTTATAGAGCCATTTTCTGAAGAA 60
DB 633 AAGCAATGTGCTTATGATAATCTCTCAGAAAGATACATTTCCATTCCTTTTTCACACTT 692
QY 61 TCGCTCCACACGACGAGTTATGACCTCAGAGTGGCGAGAGGGCTTTTGTAAAGGGGAAA 120
DB 693 CAGCTACACACACATGATGGAGAAAATTAAAGAGTTATCTGAATTATGTCTAAGTCAAAA 752
QY 121 TTAATCGACGTGGAAGGAGGAAAAGTCGTTATTCTCCAAAGGGAATAGCCCTTAATC 180
DB 753 ATCATCAACCTTTCTAATGAAGGCGAGCGCAAAAGTAGTAAAGCAAAAGGACAAGAAC 812
QY 181 CTAACCCCTCGAGAGGATAAAGTTG 204
DB 813 AATAACTTCTCAAGATAAACCTAG 836

Search completed: January 31, 2002, 10:48:51
Job time: 64 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 10:02:27 ; Search time 1174.47 Seconds
(without alignments)
4309.401 Million cell updates/sec

Title: US-08-957-709-70
Perfect score: 471
Sequence: 1 ATGCTACTCCAGACTGGAA.....CAAGAGAAAGAACTCTAG 471

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_htc: *
10: gb_estl: *
11: gb_est2: *
12: gb_htc: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	40.4	8.6	1101	13	CNS0039G
C 2	39	8.3	1101	13	BF281615
C 3	37.8	8.0	1101	13	CNS0106X
C 4	37.6	8.0	1101	13	CNS0006J
C 5	37	7.9	1019	13	CNS0107N
C 6	36.8	7.8	997	13	CNS0005TE
C 7	36.4	7.7	577	13	AQ416955
C 8	36	7.6	973	11	BG261309
C 9	35.8	7.6	1101	13	CNS000LT2
C 10	35.4	7.5	1204	13	CNS016E2
C 11	35.2	7.5	2275	10	AF034173
C 12	34.6	7.3	467	10	AI825671

13	34.6	7.3	721	11	BF278028
14	34.6	7.3	1100	13	CNS014RI
C 15	34.4	7.3	516	13	AQ571364
C 16	34.2	7.3	548	13	AQ699577
C 17	34.2	7.3	1087	13	CNS06CUN
C 18	34	7.2	518	11	BG353517
C 19	34	7.2	737	13	AZ189339
C 20	34	7.2	859	13	CNS004YY
C 21	33.8	7.2	493	13	BH029483
C 22	33.6	7.1	385	13	AQ072042
C 23	33.6	7.1	870	11	BI196947
C 24	33.6	7.1	961	13	CNS046KO
C 25	33.6	7.1	964	13	CNS0155Q
C 26	33.4	7.1	356	11	D32687
C 27	33.4	7.1	367	11	C70559
C 28	33.4	7.1	437	11	BF473989
C 29	33.4	7.1	614	13	AQ753165
C 30	33.4	7.1	648	11	BG905021
C 31	33.4	7.1	650	11	BG906626
C 32	33.4	7.1	711	13	AQ960094
C 33	33.4	7.1	1101	13	CNS017HF
C 34	33.2	7.0	518	13	AZ986009
C 35	33.2	7.0	772	13	AZ518463
C 36	33.2	7.0	856	13	AQ747571
C 37	33	7.0	399	11	BF365040
C 38	33	7.0	497	13	AQ657311
C 39	33	7.0	557	11	BG622197
C 40	33	7.0	602	11	BF992394
C 41	33	7.0	665	13	AQ693439
C 42	33	7.0	675	13	AZ336783
C 43	33	7.0	683	13	TA371H04Q
C 44	33	7.0	686	13	AZ334611
C 45	33	7.0	708	11	BG623120

ALIGNMENTS

RESULT 1
CNS0039G/c

LOCUS
DEFINITION

CNS0039G 1101 bp DNA
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly). genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

AL063921
AL063921.1 GI:4941778
GSS.
fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osada and Aaron Mannosier in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES
source
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPI-98"
/clone="BACR08K10"
/note="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 8.6% Score 40.4; DB 13; Length 1101;
Best Local Similarity 13.5%; Pred. No. 0.4;
Matches 52; Conservative 182; Mismatches 150; Indels 0; Gaps 0;

QY 21 AATCAGAAAGAAATACATATAGACCCATTTCTCAAGAAATCGCTCCACACGACGAGTTA 80
DB 1077 WTTWKRRADRDAGDADRAWDGAGTWTATWWWWWWATWDTWTKWWWWATAAK 1018

QY 81 TGACCTCAGAGTGGCAGAGAGGCTTTTGTAAAGGGGAAATTAATCGACGCTGGAAAGGA 140
DB 1017 TDTATWTTAWRADWAGRDGAGRKRDRAATDADGAGRRDGRKDKDKDKDDDDKK 958

QY 141 AGGAAAGTCGTATTCTCCTCAAGGGAATAGCCCTTAATCTCAACCTCGAGAGGATAA 200
DB 957 GKKKKAAGAATKWWDDWDWDKWDGAKRDADDGAGDKDDGKGDADDDTD 898

QY 201 GTTGCCCGACGATGTATGGGGGATATCAAGATAAGGAGCAGTTTAGCAAGAGAGGGGT 260
DB 897 GTKDDDDKDDKWDKAKGTGADTAWAATDWWGWDADMTWDAADDDWADDDRD 838

QY 261 TATTGTTCTTTTCTGCTGGTGGTACCCAGGATGGATGGAACCTTAACACTAATGCTCTA 320
DB 837 AWAKWDDAWAGARTADRDWDGDRAGRGARKRRDRKDRADDDRRDAATWTTW 778

QY 321 CAATGCCCTCAATGAACCTGCGATTAAAGATATGAGAGAGATTGCGAGATCGCATT 380
DB 777 TTTTTRDDDKWKWTDWTRWAADTWDDBDDDDDDDRAGTAGKWRRTWKRRKRRTRW 718

QY 381 TATAAGGCTAGAGGCTCGCGCAAG 404
DB 717 DDADADTARDRRRRGGDGADAG 694

RESULT 2
BF281615/c
LOCUS BF281615 336 bp mRNA EST 28-NOV-2000
DEFINITION Rattus norvegicus cDNA clone RGIAN54, mRNA sequence.
ACCESSION BF281615
VERSION BF281615.1 GI:11212685
KEYWORDS EST
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 336)
Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizi,B., Pertea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Unpublished (2000)
Other ESTs: EST446205 EST446207
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
Location/Qualifiers
1..336

FEATURES
source
Query Match 8.0% Score 37.8; DB 13; Length 1101;
Best Local Similarity 15.2%; Pred. No. 2.2;

/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RGIAN54"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/tissue_type="mixed tissue"
/lab_host="DHS-alpha"
/note="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RLI, RPL, REM, RMU, RSP
RHE, RPC, RPN"
BASE COUNT 83 a 61 c 62 g 130 t
ORIGIN

Query Match 8.3% Score 39; DB 11; Length 336;
Best Local Similarity 52.1%; Pred. No. 0.93;
Matches 87; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 90 AGTGGGACAGAGGCTTTTGTAAAGGGGAAATTAATCGACCTGGAAAGGAGGAAAAAGT 149
DB 335 AGTAGGCACTGCTGCTCTCAAAAAGGGGAAAGAGGAGGAGGAGGGGT 276

QY 150 CGTTATTCTCAAGGGAATACGCTTAATCTCAACCTCGAGAGGATAAAGTTGCCCGA 209
DB 275 CATGAGTCATGCCACCAATCATCAATAATTAATCATGTAACAGATAATTCGAAAA 216

QY 210 CGATGTTATGGGGATATGAAGATAAGGAGCAGTTTAGCAAGAGGAAG 256
DB 215 AAAATTATAGCGATAAGGAATGAAAGGATTTAGGTACAAAAG 169

RESULT 3
CNS0106X 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
fly) genomic survey sequence.
ACCESSION AL098593
VERSION AL098593
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03K20"
/note="end : T7"
BASE COUNT 258 a 107 c 60 g 175 t 501 others
ORIGIN

Query Match 8.0% Score 37.8; DB 13; Length 1101;
Best Local Similarity 15.2%; Pred. No. 2.2;

```

Matches 50; Conservative 151; Mismatches 128; Indels 0; Gaps 0;

Qy 135 AAAGGAAGAAAGTCGTTATTCTCCCAAGGAATACGCCCTTAATCTCAACCCGAGAC 194
 Db 749 AKADARRRWDKDKKKRKAARAKAADAADADAKADAKADAKADADADADGDDG 808
 Qy 195 GATAAGTTGCCGAGCATGTTATGGGGATATGAAGATAAGGAGCAGCTTTAGCAAGAGA 254
 Db 809 KKRAKDRKKKKKDKGDKKAKKAAKADAAAGDAKAKRRRRAGDKDKAKADAKA 868
 Qy 255 AGGGTTATTGGTCTTGGTCTGGGTTGACCCAGGATGGGATGGAACCTTAACACTAAT 314
 Db 869 AKKAKADDDDAKATRAKATKAKKAKAKKAKKDKDKAKAKADKDKDKDDDD 928
 Qy 315 GCTCTACAATGCTCAATGAACCTGTCGAATTAAGATATGAGAGAGATTTGTGCAGAT 374
 Db 929 KDKKDAKADKKKKDKAKKDKKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDK 988
 Qy 375 CGCATTTTAAGGCTAGAGGTCGGGCAAGAACCCCTTACAGAGGAATCTATCAGGGGAG 434
 Db 989 AKAKAADAADDAADAKAKADAKAKAKADDAKAKADAKADAKADAKAKAKRAAK 1048
 Qy 435 CACAGGTTAGCGTTTCAAGAGAAAGA 463
 Db 1049 AKADAAKAKAKADDDKDKAKDDADKA 1077

RESULT 4

CNS0006J 1101 bp DNA 03-JUN-1999

LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #

DEFINITION BACR01M22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL062049

VERSION AL062049.1 GI:4938511

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR01M22"

/note="end : TET3"

333 a 162 c 148 g 177 t 281 others

BASE COUNT

ORIGIN

Query Match 8.0%; Score 37.6; DB 13; Length 1101;

Best Local Similarity 29.4%; Pred. No. 2.6;

Matches 131; Conservative 94; Mismatches 218; Indels 3; Gaps 1;

Qy 19 AAATCAGAAAAGAAATCTATTATAGAGCCATTTTCTGAAGAATCGCTCCACACGACGAGT 78
 Db 489 AAAAARARATGARAATRWARAGRTTATTTGAAAAAATAAATTTTAAAGAAAAAAA 548
 Qy 79 TATGACCTCAGAGTGGGAGAGGCTTTTCTTAAGGGGAATTAATCGACGTGGAAG 138
 Db 549 AAATATATAAATTTTTRRTTRAGTATTTTRAKAAAAAATAAATAATATAGAAAAAG 608
 Qy 139 GAAGGAAAAGTCGTTATTCCTCCCAAGGAATACGCCCTTAATCTCAACCCGAGAGGATA 198
 Db 609 AAAAAAARAAGTGTGTTGAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 668
 Qy 199 AAGTTGCCGAGCATGTTATGGGGATATGAAGATAAGGAGCAGCTTTAGCAAGAGAGG 258
 Db 669 AATAAAAAAARATGRTATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 728
 Qy 259 GTTATTGCTTCTTTGCTTGGTGGTGCACCCAGGATGGGATGGAACCTTAACACTAATGCTC 318
 Db 729 AWRRAAAATKKKKTKTKRRAAGRR---ARRAWAGAAAAAARAADAATKTGKTWAKA 785
 Qy 319 TACAATGCTCAATGAACCTGTGCAATTAAGATATGAGAGAGAGATTTGTGCAGATCGCA 378
 Db 786 AWRTAANKDKWKATDAAAKAAARTRDWTTAKAKDKTKWGAATAATGAGAGAARRRWR 845
 Qy 379 TTTATAAGGCTAGAGGTCGGCAAGAAACCCCTTACAGAGGAACCTATCAGGGGAGCACA 438
 Db 846 KGKDGTRARRARGAGDGDWDAKAWAWAATAAATTAARWDTATADDDRRAKAWDKRAAAA 905
 Qy 439 AGGTTAGCGTTTCAAGAGAAAGAA 464
 Db 906 RDKDGRAGGTGWRRAARARTRKA 931

RESULT 5

CNS0107N/c 1019 bp DNA 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC

DEFINITION BACN03D10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL098621

VERSION AL098621.1 GI:5610232

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Plasmid Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1019)

TITLE Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES

source

1. .1019

/organism="Drosophila melanogaster"

/plasmid="pBelOBAC11"

/db_xref="taxon:7227"

/clone_lib="DrosBAC"

/clone="BACN03D10"

/note="end : SP6"

BASE COUNT 226 a 163 c 104 g 400 t 126 others

Query Match 7.9% Score 37; DB 13; Length 1019;
Best Local Similarity 41.8% Pred. No. 3.8; Mismatches 0; Gaps 0;
Matches 97; Conservative 10; Indels 125; Indels 0; Gaps 0;

Qy 233 TAAGGAGCAGTTACGACAGAGAGGGTANTGGTTCTTTTGGCTGGGTGACCCAGAT 292
Db 623 TAAATWATAAACAACAGATAGMGGTTTAAAGAAATTTATGGAAATTTTAAAGAAGT 564
Qy 293 GGGATGGAATTAACACTAATGCTCTACATGCCTCAATGACCTTCGAATTAAGAT 352
Db 563 AGAATGAATGGAATTAATAAAGAGAGGAATAATATAAAGTAAAWAATAAGAG 504
Qy 353 ATGGAGAGAGATTTGTCAGATCGCATTTATAAGCTTAGAGGGTCCGGCAGAAACCCCT 412
Db 503 AWATGAAATWAGAAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 444
Qy 413 ACAGAGGAATCTCAGGGGAGCACAAGGTTAGCGTTTTCRAAGAGAGAAAGAA 464
Db 443 ANAAGAWAAATAAGAGAAAGAAAGAAATGAGAAAAAAGTGAAGGA 392

RESULT 6
CNS005TE/c 997 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BAC12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL060767.1 GI:4943573
VERSION 1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> the BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammot in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

1. 997
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BAC12K22"
/note="end : TET3"
/note="end : TET3"
89 a 99 c 13 g 258 t 538 others

FEATURES
source

Query Match 7.8% Score 36.8; DB 13; Length 997;
Best Local Similarity 17.2% Pred. No. 4.3;

Matches 78; Conservative 154; Mismatches 222; Indels 0; Gaps 0;
Qy 12 AGACTGGAAATCAGAAAGAAATACATATAGAGCCATTTTCTGAAGATCGCTCAACC 71
Db 931 AAARERARRRAGRAGRAGRSAGGGRGRRRAAARARARARARARRR 872
Qy 72 ACCAGGTTATCACCTACAGTGGCAGAGAGCTTTTGTAAAGGGAAATTAATCGAGT 131
Db 871 AKAARGARRRRRRRRRRAGRRRAGRRRGRGRRRARRRARRRARRRARRR 812
Qy 132 GAAAGGAAGAAAGTCGTTATTCCTCAAGGAATACGCCTTAATCTTAACCTCGA 191
Db 811 GRRRAAARARRRRRRRRRAGRRRRRGRGRRRRRRRRRARRRARRRARRRARRR 752
Qy 192 GAGGATAAGTTGCCCGACGATGTTATGGGGGATATGAAGATAAGGACGACTTTAGCAAG 251
Db 751 GARAAARRRRARRAGAAARRRRRRRRRAGARRRRRRRRRRARRRARRRARRRARRR 692
Qy 252 AGAAGGGTTATGTTCTTTTGGTTGGTTCACCCAGGATGGGATGGAACCTTAACACT 311
Db 691 FRRGAGARRRRRRRRRAGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 632
Qy 312 AATGCTCTACAAATGCCTCAATGAACCTGTCTGAATTAAGATATGAGAGAGATTTGTGCA 371
Db 631 RRR 572
Qy 372 GATCCCATTTTAAAGGCTAGAGGTCGCGCAAGAACCCCTTACAGAGAGAACTACAGGG 431
Db 571 RGRARRAAGAAARR 512
Qy 432 GAGCACAGGTTAGCGTTTTCACAAAGAGAGAGAAA 465
Db 511 RGGARR 478

RESULT 7
A0416995/c 577 bp DNA GSS 23-MAR-1999
LOCUS RPCI-11-179G24.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-179G24
DEFINITION DNA sequence.
ACCESSION A0416995
VERSION 1 GI:4475193
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 577)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
Unpublished (1997)
Other_GSSs: RPCI-11-179G24.TV

CONTACT: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6
Class: BAC ends.

Location/Qualifiers
1. 577
/organism="Homo sapiens"
/db_xref="GDB:7568519"

FEATURES
source

/db_xref="taxon:9606"
 /clone="RPCI-11-179G24"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPCI11 Human Male BAC Library"
 BASE COUNT 157 a 110 c 139 g 169 t 2 others
 ORIGIN

Query Match 7.7%; Score 36.4; DB 13; Length 577;
 Best Local Similarity 49.0%; Pred. No. 5.4;
 Matches 123; Conservative 0; Mismatches 127; Indels 1; Gaps 1;
 QY 4 CTACTTCCAGCTGGAATAATCAGAAAGAAATACCTTTATAGAGCCATTTCTGAAGATCG 63
 DB 520 CTAGTTTTCATGCTATGCTCAGGAGAACTTCTGTACCAACATGCATCTTCCCATCA 461
 QY 64 CTCACACGAGCAGGTTATGACCTCAGAGTGGGAGAGAGGCTTTTGTAAAGGGGAATTA 123
 DB 460 CCAGACCTAGATAATGCTATGCTTGAATGCTGAAGAGGATNAGATGAAGATCAAA 401
 QY 124 ATGAGCTGGAAGAAGGAAGTGGTATTCCTC-CAAGGGAATACGCTTAACTCT 182
 DB 400 GACAATGGGGAGGAAAGAAACATAGTTTAAACACACCATGGGAGACTATATATGCC 341
 QY 183 AACCTCGAGAGGTAAGTGGCCGAGCATGTTATGGGGATATGAAGATAGGAGCAG 242
 DB 340 ACCCGAAGAAATCTATAAGCTGAAGTACATGACTGTGAACAGAAGGCAATGAAGCAG 281
 QY 243 TTTAGCAAGAG 253
 DB 280 ATAGAGAGAG 270

RESULT 8
 BG261309
 LOCUS 602373061F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:448463 5',
 DEFINITION mRNA sequence.
 ACCESSION BG261309
 VERSION BG261309.1 GI:12771125
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 973)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: AFCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM10324 row: j column: 16
 High quality sequence stop: 705.

FEATURES
 source
 1..973
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:448463"
 /clone_lib="NIH_MGC_93"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed."

Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 BASE COUNT 318 a 172 c 230 g 253 t
 ORIGIN

Query Match 7.6%; Score 36; DB 11; Length 973;
 Best Local Similarity 50.6%; Pred. No. 7.3;
 Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
 QY 286 CCAGGATGGGATGAAACTTAACACTAATGCTTACAAATGCTCAATGAACCTGTGCAA 345
 DB 708 CCACGACTTATTGTAGCGGTGACACTTGACAAGTTCACAAATTCCTGTGCAA 767
 QY 346 TTAGATATGGAGAGAGATTTGTCAGATCCCATTTATAGGCTAGAGGTCGCGCAGA 405
 DB 768 TTTTGATAGTACGACAAAGCATGTTTCGTCGCAAAAAGGCTTACGTTCTCCGAAACA 827
 QY 406 AACCTTACAGAGAAACTATCAGGGGAGCAGCAAGGTTAGCGTTTTCAAAGA 457
 DB 828 ATCTTTATATCTAGCGGTATATAAAGTACACGGTAAACGGCGTCCACA 879

RESULT 9
 CNS00LT2
 LOCUS 1101 bp DNA GSS 14-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
 BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL078714
 VERSION AL078714.1 GI:5102004
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR48P19"
 /note="end : TET3"
 BASE COUNT 469 a 6 c 69 g 151 t 406 others
 ORIGIN

Query Match 7.6%; Score 35.8; DB 13; Length 1101;
 Best Local Similarity 17.6%; Pred. No. 8.5;
 Matches 60; Conservative 128; Mismatches 153; Indels 0; Gaps 0;


```
VERSION AL104520.1 GI:5616534
KEYWORDS
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES             Location/Qualifiers
     source          1..1100
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelobAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN12G24"
                     /note="end : SP6"
BASE COUNT   184 a   175 c   118 g   190 t   433 others
ORIGIN
Query Match       7.3%; Score 34.6; DB 13; Length 1100;
Best Local Similarity 7.7%; Pred. No. 19;
Matches 24; Conservative 91; Mismatches: 195; Indels 0; Gaps 0;
QY 68 AACGACGAGGTATGACCTCAGAGTGGCAGAGCGCTTTGTTAAAGGGGAAATTAATCG 127
Db 669 AACNTNNNNNNKKNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNA 728
QY 128 ACGTGGAAGAGGAGGAAAGTCTTATCTCCAGGGATACGCCTTAATCCTAACCC 187
Db 729 NNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 788
QY 188 TCGAGAGGATAAAGTTCGCCGACCATGTTATGGGGGATATGAAGATAGGAGCAGTTAG 247
Db 789 TKNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 848
QY 248 CAAGAGAGGGGTTATGGTCTTTCTGCTGGGTTGACCCAGGATGGATCGAATCTAA 307
Db 849 NNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 908
QY 308 CACTAATGCTCTACAATGCCTCAATCAACTGTCGAATTAAGATATGGAGAGAGATTG 367
Db 909 KKKKVKKKKKMMKTTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTMTT 968
QY 368 TGCAGATCGC 377
Db 969 KKKCKMKCKC 978

RESULT 15
AQ571364/c
LOCUS
DEFINITION HS_5376_A2_D03_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=952 Col=6 Row=G, DNA sequence.
ACCESSION AQ571364
VERSION AQ571364.1 GI:4964584
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 516)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL PROC. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 952 row: G column: 6
Seq primer: T7
Class: BAC ends
High quality sequence stop: 516.
Location/Qualifiers
     source          1..516
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="plate-952 Col=6 Row=G"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACE3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"
BASE COUNT   153 a   93 c   84 g   177 t   9 others
ORIGIN
Query Match       7.3%; Score 34.4; DB 13; Length 516;
Best Local Similarity 52.6%; Pred. No. 20;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 198 AAAGTTCGCCGACGATCTTATGGGGATATGAAGATAGGAGCAGTTAGCAAGAGAGG 257
Db 494 AACTGACACCAACATATGCTGTGAGGATCTGGAGANACAGGAACCTCTATTTTTTNT 435
QY 258 GGTATTGTTCTTTCTGTTGGTTGACCCAGGATGGGATGGAACCTAACACATAGCT 317
Db 434 TTTTNTTTTTTTTTTCTGTTGGGATTCAAATGATATGAAACCTAGNACCACAGTT 375
QY 318 CTACAATGCCTCAA 332
Db 374 TGACAGTTCTTTAGA.360

Search completed: January 31, 2002, 10:22:30
Job time: 1203 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:15 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-81

Perfect score: 14

Sequence: 1 TGLIDPGFGELKL 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	19	AAW72857
2	6	42.9	152	22	AAAG23390
3	6	42.9	159	21	AAAG19393
4	6	42.9	177	13	AAAR20194
5	6	42.9	177	21	AAAV78000
6	6	42.9	177	21	AAAV51092
7	6	42.9	219	21	AAAG19395
8	6	42.9	236	21	AAAG23392
9	6	42.9	242	21	AAAG23391
10	6	42.9	249	21	AAAG19394
11	6	42.9	249	21	AAAG53467

12	6	42.9	255	21	AAAG23390	Arabidopsis thalia
13	6	42.9	256	22	AAE02484	Arabidopsis thalia
14	6	42.9	268	21	AAAG19393	Arabidopsis thalia
15	6	42.9	288	21	AAAG53466	Arabidopsis thalia
16	6	42.9	284	21	AAAG05943	Protein deduced fr
17	6	42.9	338	21	AAAG53465	Arabidopsis thalia
18	6	42.9	346	22	AAAB36217	Murine immune asso
19	6	42.9	406	21	AAAB58169	Lung cancer associ
20	6	42.9	429	21	AAAG41644	Arabidopsis thalia
21	6	42.9	436	21	AAAG41643	Arabidopsis thalia
22	6	42.9	447	21	AAAG41642	Arabidopsis thalia
23	6	42.9	1669	22	AAAM40863	Human polypeptide
24	6	42.9	1672	22	AAAM39077	Human polypeptide
25	5	35.7	10	22	AAAG87390	Saccharomyces cere
26	5	35.7	10	22	AAAG87391	Saccharomyces cere
27	5	35.7	17	13	AAAR28307	N-terminal of lysl
28	5	35.7	28	20	AAAW99784	Rat AIM-1 peptide
29	5	35.7	29	19	AAAY21188	Human bcl2 proto-o
30	5	35.7	31	18	AAAW30963	Timothy grass poll
31	5	35.7	31	18	AAAW30963	Timothy grass poll
32	5	35.7	35	18	AAAW30931	Birch pollen aller
33	5	35.7	36	21	AAAG12154	Arabidopsis thalia
34	5	35.7	38	15	AAAR58378	TSAR binding domai
35	5	35.7	39	18	AAAW31032	Mugwort pollen all
36	5	35.7	39	18	AAAW31065	Mugwort pollen all
37	5	35.7	40	22	AAAM15404	Peptide #1838 enco
38	5	35.7	40	22	AAAM27892	Peptide #1929 enco
39	5	35.7	40	22	AAAM03165	Peptide #1847 enco
40	5	35.7	44	21	AAAB32423	Human secreted pro
41	5	35.7	52	22	AAAU01986	Gene #23 human sec
42	5	35.7	54	20	AAAY42491	Collagen type I-de
43	5	35.7	59	21	AAAG59623	Arabidopsis thalia
44	5	35.7	62	22	AAE02716	Human alpha (I) t
45	5	35.7	62	22	AAAB68070	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAW72857
ID AAW72857 standard; Peptide; 14 AA.
XX
AC AAW72857;
XX
DT 01-MAR-1999 (first entry)
XX
DE Herpesvirus dUTPase uridine-binding motif.
DE
DE Polymerase enhancing factor; PEF; dUTPase; PCR; amplification;
KW sequencing; replication.
KW
XX
OS Herpesvirus.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX
XX (STRA-) STRATAGENE.
XX
XX Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes -
PT Improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication

PS Claim 71; Page 47; 161pp; English.

XX This is the uridine-binding motif of herpesvirus dnpase. Sequences

CC are provided (see AAW72849-57) of the uridine-binding motifs of the

CC dnpases and dCpP deaminases of Pyrococcus furiosus (see also

CC AAW72847), Methanococcus jannaschii, Desulfurolobus ambivalens,

CC Escherichia coli, yeast, human and herpesvirus; a consensus (see

CC AAW72848) is also provided. A claimed method of enhancing a nucleic

CC acid polymerase reaction comprises performing the reaction in the

CC presence of one or more of the following: a polymerase enhancing

CC factor (PEF), a dnpase, a protein that turns-over dUTP and a

CC protein having one or more of the sequences provided in AAW72848-57.

CC A claimed protein having PEF activity comprises one or more of

CC sequences given in AAW72848-57. Kits are provided for replicating

CC nucleic acids, for site-directed mutagenesis, for nucleic acid

CC sequencing or for amplification (preferably PCR or RT-PCR).

XX

SQ Sequence 14 AA;

Query Match 100.0%; Score 14; DB 19; Length 14;

Best Local Similarity 100.0%; Pred. No. 1.4e-08;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDPGFGQGLKL 14

Db 1 tglidpgfgqglkl 14

RESULT 2

AAAG90573

ID AAG90573 standard; Protein; 152 AA.

XX

AC AAG90573;

XX

DT 26-SEP-2001 (first entry)

XX

DE C glutamicum protein fragment SEQ ID NO: 4327.

XX

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

XX

OS Corynebacterium glutamicum.

XX

PN EPL108790-A2.

XX

PD 20-JUN-2001.

XX

PF 18-DEC-2000; 2000EP-0127688.

XX

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX

WPI; 2001-376931/40.

DR N-PSDB; AAH65792.

XX

PT Novel polynucleotides derived from Corynebacterium, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT

XX

Claim 17; SEQ ID NO: 4327; 246pp + Sequence Listing; English.

PS

XX The present invention provides a number of nucleotide and protein

CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of corynebacterium bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Corynebacterium bacterium, and identifying a homologue of a gene derived

CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX

SQ Sequence 152 AA;

Query Match 42.9%; Score 6; DB 22; Length 152;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDPGFQ 9

Db 145 idpgfq 150

RESULT 3

AAAB24671

ID AAB24671 standard; Peptide; 159 AA.

XX

AC AAB24671;

XX

DT 27-NOV-2000 (first entry)

XX

DE Plant SDF encoded polypeptide sequence SEQ List 1 NO:95.

XX

KW Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;

KW SDF; genetic mapping; identification; promoter; structural gene; UTR;

KW untranslated region; expression control.

XX

OS Plant.

XX

PN WO2000040695-A2.

XX

PD 13-JUL-2000.

XX

PF 07-JAN-2000; 2000WO-US00466.

XX

PR 08-JAN-1999; 99US-0115293.

XX

PA (CERE-) CERES INC.

XX

PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;

PI Zheng L;

XX

WPI; 2000-465970/40.

XX

PT New corn plant and Arabidopsis thaliana sequence-determined DNA

PT fragments, useful for expressing gene products and for controlling

PT expression of a target gene -

XX

PS Claim 14; Page 369-370; 673pp; English.

XX

CC The present invention describes polynucleotides, such as complete cDNA

CC sequences and/or sequences of genomic DNA encompassing complete genes,

CC portions of genes, and/or intergenic regions, collectively referred to

CC as sequence-determined DNA fragments (SDFs), from corn plants and

CC Arabidopsis thaliana. The SDFs are promoters, structural genes,

CC untranslated regions (UTRs), or 3' termination sequences. They can be

CC used for expressing a gene product and controlling expression of a

CC target gene, either as a promoter, a structural gene, an UTR or as a

CC 3' termination sequence. They are also useful as tools for genetic

CC mapping, and identification of a particular individual plant or for

CC clustering a group of plants with a common trait. AAA78433 to AAA78630

CC and AAB24605 to AAB25099 represent the specifically claimed

CC polynucleotide sequences and polypeptides encoded by them given in the

CC present invention.

```

XX SQ Sequence 159 AA;
Query Match 42.9%; Score 6; DB 21; Length 159;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDP 6
Db 75 tglidp 80
|||||

RESULT 4
AAR20194
ID AAR20194 standard; Protein; 177 AA.
XX AC AAR20194;
XX DT 15-APR-1992 (first entry)
XX DE Fungal alpha-sarcine.
XX KW alpha-hypoxanthine; signal sequence; fungus.
XX OS Aspergillus giganteus.
XX FH Key Location/Qualifiers
FT Peptide 1..27
FT Protein /label= Signal
FT /label= alpha_sarcine
XX JF03266986-A.
XX DT 27-NOV-1991.
XX PF 14-SEP-1990; 90JP-0244181.
XX PR 22-FEB-1990; 90JP-0044112.
XX PR 14-SEP-1990; 90JP-0244181.
XX (SAKA ) OTSUKA SEIYAKU KOJY.
XX WP1: 1992-019323/03.
XX N-PSDB; AAQ20385.
PT Alpha-sarcine gene - allows large scale efficient prepn. using
PT e.g. Aspergillus sp. microorganism
XX PS Disclosure; Page 3; 12pp; Japanese.
XX CC The alpha-sarcine coding sequence was isolated from a cDNA
CC expression library following screening with an antibody. The amino
CC acid sequence was deduced from the cDNA sequence..
XX SQ Sequence 177 AA;

Query Match 42.9%; Score 6; DB 13; Length 177;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QGELKL 14
Db 169 qgelkl 174
|||||

RESULT 5
AA178000
ID AA178000 standard; protein; 177 AA.
XX AC AA178000;

```

```

XX DT 23-JUN-2000 (first entry)
XX DE Amino acid sequence of RNAS_ASFGI.
XX KW Functional site descriptor; FSD; enzyme; protein structure;
XX three-dimnsional structure; active site.
XX OS Unidentified.
XX PN WO200011206-A1.
XX PD 02-MAR-2000.
XX PF 27-MAY-1999; 99WO-US11913.
XX PR 25-AUG-1998; 98US-0099300.
XX PR 16-FEB-1999; 99US-0120311.
XX (SCRI ) SCRIPPS RES INST.
XX PI Skolnick J, Fetrov JS;
XX WP1: 2000-224711/19.
XX Functional site descriptor for proteins, useful for predicting protein
XX activity -
XX Example 1; Fig 12; 172pp; English.
XX The invention provides a functional site descriptor (FSD) that defines a
XX spatial configuration of a functional site (other than a divalent metal
XX ion binding site) in a protein (I), for use in an inexact three-
XX dimensional (3D) structure model of (I) to determine if (I) possesses
XX the function associated with FSD. FSD is a set of geometrical constraints
XX for at least one atom in each of at least two amino acid (aa) residues
XX that comprise the functional site of (I) and at least one aa is
XX identified as a particular aa (or set of aa). The atoms are: (i) amide
XX nitrogen, alpha or carbonyl carbon, or carbonyl oxygen within the polymer
XX backbone; (ii) a beta-carbon of aa, or (iii) a pseudatom, with at least
XX one being of type (i). FSD are used to predict biological activities of
XX proteins, e.g. enzymatic, ligand-binding or protein interaction
XX activities, of animal, plant, prokaryotic or viral origin. They may also
XX be used to validate/confirm results from other predictive methods. The
XX method can be used even with inexact (computed) models of protein
XX structure, deduced from protein or nucleic acid sequence data, i.e. an
XX accurate 3D model of the whole protein is not required, rather just the
XX 3D orientation of the active site. Sequences AAY77991-Y78000 and
XX AAY77923-924 represent peptides used in the method of the invention.
XX SQ Sequence 177 AA;

Query Match 42.9%; Score 6; DB 21; Length 177;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QGELKL 14
Db 169 qgelkl 174
|||||

RESULT 6
AA51092
ID AA51092 standard; protein; 177 AA.
XX AC AA51092;
XX DT 22-MAR-2000 (first entry)
XX DE Functional site descriptor protein RNAS_ASFGI.
XX KW Functional site descriptor; FSD; 3-dimensional; 3-D; structural model;

```


PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 25-JUN-1999; 99US-0140354.
PR 26-JUN-1999; 99US-0140354.
PR 27-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140354.
PR 29-JUN-1999; 99US-0140354.
PR 30-JUN-1999; 99US-0140354.
PR 01-JUL-1999; 99US-0141287.
PR 02-JUL-1999; 99US-0141287.
PR 03-JUL-1999; 99US-0141287.
PR 04-JUL-1999; 99US-0141287.
PR 05-JUL-1999; 99US-0141287.
PR 06-JUL-1999; 99US-0141287.
PR 07-JUL-1999; 99US-0141287.
PR 08-JUL-1999; 99US-0141287.
PR 09-JUL-1999; 99US-0141287.
PR 10-JUL-1999; 99US-0141287.
PR 11-JUL-1999; 99US-0141287.
PR 12-JUL-1999; 99US-0141287.
PR 13-JUL-1999; 99US-0141287.
PR 14-JUL-1999; 99US-0141287.
PR 15-JUL-1999; 99US-0141287.
PR 16-JUL-1999; 99US-0141287.
PR 17-JUL-1999; 99US-0141287.
PR 18-JUL-1999; 99US-0141287.
PR 19-JUL-1999; 99US-0141287.
PR 20-JUL-1999; 99US-0141287.
PR 21-JUL-1999; 99US-0141287.
PR 22-JUL-1999; 99US-0141287.
PR 23-JUL-1999; 99US-0141287.
PR 24-JUL-1999; 99US-0141287.
PR 25-JUL-1999; 99US-0141287.
PR 26-JUL-1999; 99US-0141287.
PR 27-JUL-1999; 99US-0141287.
PR 28-JUL-1999; 99US-0141287.
PR 29-JUL-1999; 99US-0141287.
PR 30-JUL-1999; 99US-0141287.
PR 31-JUL-1999; 99US-0141287.
PR 01-AUG-1999; 99US-0141287.
PR 02-AUG-1999; 99US-0141287.
PR 03-AUG-1999; 99US-0141287.
PR 04-AUG-1999; 99US-0141287.
PR 05-AUG-1999; 99US-0141287.
PR 06-AUG-1999; 99US-0141287.
PR 07-AUG-1999; 99US-0141287.
PR 08-AUG-1999; 99US-0141287.
PR 09-AUG-1999; 99US-0141287.
PR 10-AUG-1999; 99US-0141287.
PR 11-AUG-1999; 99US-0141287.
PR 12-AUG-1999; 99US-0141287.
PR 13-AUG-1999; 99US-0141287.
PR 14-AUG-1999; 99US-0141287.
PR 15-AUG-1999; 99US-0141287.
PR 16-AUG-1999; 99US-0141287.
PR 17-AUG-1999; 99US-0141287.
PR 18-AUG-1999; 99US-0141287.
PR 19-AUG-1999; 99US-0141287.
PR 20-AUG-1999; 99US-0141287.
PR 21-AUG-1999; 99US-0141287.
PR 22-AUG-1999; 99US-0141287.
PR 23-AUG-1999; 99US-0141287.
PR 24-AUG-1999; 99US-0141287.
PR 25-AUG-1999; 99US-0141287.
PR 26-AUG-1999; 99US-0141287.
PR 27-AUG-1999; 99US-0141287.

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 219;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 3 LIDPGF 8
Db 86 lidpgf 91

RESULT 8
AAG23392
ID AAG23392 standard; Protein; 236 AA.

XX
AC AAG23392;
XX
DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 26681.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX Hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

OS Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX XX 25-FEB-2000; 2000EP-0301439.
XX XX 25-FEB-1999; 99US-0121025.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131089.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 23-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149920.
PR 25-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 236;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFOGEL 12

Db 188 gfggel 193

RESULT 9

AAAG23391
 ID AAAG23391 standard; Protein; 242 AA.

XX AC AAAG23391;

XX DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 26680.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 17-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147004.
PR 04-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147202.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 08-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147535.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 15-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155559.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161392.
PR 28-OCT-1999; 99US-0161593.
PR 28-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 242;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GPOGEL 12

Db 194 gfggel 199

RESULT 10

AAG19394
ID AAG19394 standard; Protein: 249 AA.

XX AC AAG19394;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21178.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 23-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9% Score 6; DB 21; Length 249;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDPGF 8
Db 116 lidpgf 121

RESULT 11
AAG53467
ID AAG53467 standard; Protein; 249 AA.
XX
AC AAG53467;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 68076.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126268.
PR 29-MAR-1999; 99US-0126785.
PR 02-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135121.
PR 21-MAY-1999; 99US-0135253.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141287.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142893.
PR 08-JUL-1999; 99US-0142893.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143342.
PR 15-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
```

```
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 249;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDPGF 8
Db 116 lidpgf 121

RESULT 12
AAG23390
ID AAG23390 standard; Protein; 255 AA.
XX
AC AAG23390;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 26679.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SRP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
```

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 18-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 18-JUL-1999; 99US-0144331.
PR 18-JUL-1999; 99US-0144332.
PR 18-JUL-1999; 99US-0144333.
PR 18-JUL-1999; 99US-0144334.
PR 18-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156459.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159233.
PR 13-OCT-1999; 99US-0159234.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

PS Claim 4; Page 99-100; 116pp; English.

XX The present amino acid sequence is Arabidopsis thaliana

CC transcription factor homologue. This novel transcription factor

CC is useful for modifying a plant's phenotype in desirable ways, such as

CC modifying a plants environmental stress. The transcription factor is

CC encoded by environmental stress tolerance gene derived from

CC Arabidopsis thaliana. The transcription factors and the genes encoding

CC them are used to alter the structure and developmental characteristics

CC of plants such as soybean, wheat, corn, potato, cotton, rice, oilseed

CC rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, papaya,

CC strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, onion,

CC cucumber, eggplant, grapes, honey dew, lettuce, mango, melon, blueberry,

CC peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato,

CC watermelon, rosaceous fruits and/or vegetable brassicas. These sequences

CC are also used for modifying traits associated with environmental stress

CC tolerance, such as freezing, chilling, heat, drought, water saturation,

CC salt, photoconditions, radiation and ozone. The transcription factors

CC are used in gene therapy.

XX SQ Sequence 256 AA;

Query Match 42.9%; Score 6; DB 22; Length 256;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GFGGEL 12

Db 207 gfggel 212

RESULT 13

AAE02484

ID AAE02484 standard; Protein; 256 AA.

XX AC AAE02484;

XX DT 10-AUG-2001 (first entry)

XX DE Arabidopsis thaliana transcription factor G526 homologue, G765.

XX KW Transcription factor; environmental stress tolerance; gene therapy;

XX KW plant structure; plant development.

XX OS Arabidopsis thaliana.

XX FH Key Location/Qualifiers

FT Domain 23..167

FT /note= "Conserved domain"

XX WO200136598-A1.

XX PN 25-MAY-2001.

XX PD 14-NOV-2000; 2000WO-US31458.

XX PF 17-NOV-1999; 99US-0166228.

XX PR 17-APR-2000; 2000US-0197899.

XX PR 22-AUG-2000; 2000US-0227439.

XX (MEND-) MENDEL BIOTECHNOLOGY INC.

PA (PINE/) PINEDA O.

PA (YUGG/) YU G.

PA (CREE/) CREELMAN R.

PA (RIEC/) RIECHMANN J L.

PA (HEAR/) HEARD J.

PA (RATC/) RATCLIFFE O.

PA (REUB/) REUBER L.

PA (KEDD/) KEDDIE J.

XX Pineda O, Yu G, Creelman R, Riechmann JL, Heard J, Ratcliffe O;

PI Reuber L, Keddie J;

PI WPI: 2001-336000/35.

DR N-PSDB; AAD06464.

XX Nucleic acids encoding plant transcription factor polypeptides, useful

PT for altering the environmental stress tolerance characteristics of

PT plants -

XX

PS Claim 4; Page 99-100; 116pp; English.

XX The present amino acid sequence is Arabidopsis thaliana

CC transcription factor homologue. This novel transcription factor

CC is useful for modifying a plant's phenotype in desirable ways, such as

CC modifying a plants environmental stress. The transcription factor is

CC encoded by environmental stress tolerance gene derived from

CC Arabidopsis thaliana. The transcription factors and the genes encoding

CC them are used to alter the structure and developmental characteristics

CC of plants such as soybean, wheat, corn, potato, cotton, rice, oilseed

CC rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, papaya,

CC strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, onion,

CC cucumber, eggplant, grapes, honey dew, lettuce, mango, melon, blueberry,

CC peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato,

CC watermelon, rosaceous fruits and/or vegetable brassicas. These sequences

CC are also used for modifying traits associated with environmental stress

CC tolerance, such as freezing, chilling, heat, drought, water saturation,

CC salt, photoconditions, radiation and ozone. The transcription factors

CC are used in gene therapy.

XX SQ Sequence 256 AA;

Query Match 42.9%; Score 6; DB 22; Length 256;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LIDPGF 8

Db 205 lidpgf 210

RESULT 14

AAG19393

ID AAG19393 standard; Protein; 268 AA.

XX AC AAG19393;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21177.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A1.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134236.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	28-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	21-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147300.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139119.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139453.	PR	11-AUG-1999;	99US-0148319.
PR	17-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	25-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139763.	PR	26-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0140353.	PR	30-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	99US-0140354.	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140695.	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140823.	PR	07-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	99US-0140991.	PR	10-SEP-1999;	99US-0153070.
PR	30-JUN-1999;	99US-0141287.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141842.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0142154.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142330.	PR	23-SEP-1999;	99US-0155139.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142970.	PR	24-SEP-1999;	99US-0156559.
PR	12-JUL-1999;	99US-0143542.	PR	28-SEP-1999;	99US-0156598.
PR	13-JUL-1999;	99US-0143644.	PR	29-SEP-1999;	99US-0157177.
PR	14-JUL-1999;	99US-0144005.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144085.	PR	06-OCT-1999;	99US-01

```
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 268;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDPGF 8
Db 135 lidpgf 140
|||||

RESULT 15
AAG53466
ID AAG53466 standard; Protein; 268 AA.
XX AC AAG53466;
XX DT
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 68075.
XX
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 18-MAY-1999; 99US-0134941.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 19-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
XX 19-JUL-1999; 99US-0144335.
XX 20-JUL-1999; 99US-0144352.
XX 20-JUL-1999; 99US-0144632.
XX 20-JUL-1999; 99US-0144884.
XX 21-JUL-1999; 99US-0144814.
XX 21-JUL-1999; 99US-0145086.
XX 21-JUL-1999; 99US-0145088.
XX 22-JUL-1999; 99US-0145087.
XX 22-JUL-1999; 99US-0145089.
XX 22-JUL-1999; 99US-0145192.
XX 23-JUL-1999; 99US-0145145.
XX 23-JUL-1999; 99US-0145218.
XX 23-JUL-1999; 99US-0145224.
XX 26-JUL-1999; 99US-0145276.
XX 27-JUL-1999; 99US-0145913.
```

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149302.
PR 23-AUG-1999; 99US-0149330.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 268;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDPGF 8
 |||||
Db 135 lidpgf 140

Search completed: January 31, 2002, 13:18:16
Job time: 177 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:16 ; Search time 65.13 Seconds
(without alignments)
4.837 Million cell updates/sec

Title: US-08-957-709-81
Perfect score: 14
Sequence: 1 TGLIDPGFGELKL 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	42.9	150	1	US-07-912-740D-48
2	6	42.9	283	4	US-09-588-256-8
3	6	42.9	346	4	US-09-049-672A-28
4	5	35.7	10	3	US-08-159-339A-761
5	5	35.7	10	3	US-08-159-339A-762
6	5	35.7	20	2	US-08-564-972-46
7	5	35.7	22	4	US-09-219-849-13
8	5	35.7	38	1	US-08-471-052A-36
9	5	35.7	38	1	US-08-189-331-36
10	5	35.7	38	2	US-08-471-939-36
11	5	35.7	38	2	US-08-471-800-36
12	5	35.7	38	2	US-08-471-068-36
13	5	35.7	67	1	US-08-464-531-86
14	5	35.7	67	2	US-08-461-598-86
15	5	35.7	67	3	US-08-322-137-86
16	5	35.7	67	3	US-09-230-944-4
17	5	35.7	141	4	US-08-364-972-63
18	5	35.7	150	2	US-08-564-972-57
19	5	35.7	180	2	US-08-052-681-1
20	5	35.7	198	1	US-09-625-188-18
21	5	35.7	230	4	US-09-237-543-4
22	5	35.7	254	4	US-08-154-915-2
23	5	35.7	292	1	US-08-464-517-38
24	5	35.7	292	2	US-08-246-361A-38
25	5	35.7	292	2	US-08-463-772-38
26	5	35.7	292	3	PCT-US93-09945-2
27	5	35.7	292	5	PCT-US93-09945-2

28	5	35.7	293	3	US-09-188-579-82	Sequence 82, Appl
29	5	35.7	293	4	US-09-315-444-82	Sequence 82, Appl
30	5	35.7	313	2	US-08-463-081B-26	Sequence 26, Appl
31	5	35.7	313	2	US-08-461-379A-26	Sequence 26, Appl
32	5	35.7	313	2	US-08-462-390B-26	Sequence 26, Appl
33	5	35.7	313	3	US-08-463-074B-26	Sequence 26, Appl
34	5	35.7	313	3	US-08-465-585C-26	Sequence 26, Appl
35	5	35.7	313	3	US-08-652-446-26	Sequence 8, Appl
36	5	35.7	313	4	US-09-237-543-9	Sequence 9, Appl
37	5	35.7	313	4	US-09-237-543-6	Sequence 6, Appl
38	5	35.7	326	4	US-09-237-543-2	Sequence 2, Appl
39	5	35.7	331	2	US-08-564-972-9	Sequence 9, Appl
40	5	35.7	359	4	US-09-179-558-65	Sequence 65, Appl
41	5	35.7	374	3	US-08-878-801-2	Sequence 2, Appl
42	5	35.7	387	1	US-08-123-161A-10	Sequence 10, Appl
43	5	35.7	387	1	US-08-493-278-10	Sequence 10, Appl
44	5	35.7	389	6	5447867-4	Patent No. 5447867
45	5	35.7	389	6	5447867-4	

ALIGNMENTS

RESULT 1
US-07-912-740D-48
; Sequence 48, Application US/07912740D
; Patent No. 5439815
; GENERAL INFORMATION:
; APPLICANT: FITTON, John E.
; APPLICANT: TIMMS, David
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHVE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/912,740D
; FILING DATE: 13-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1586-2
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-912-740D-48

Query Match 42.9%; Score 6; DB 1; Length 150;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QGELKL 14

Db 142 QGELKL 147

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-761

Query Match 35.7%; Score 5; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFOQE 11
|||||
DB 5 GFOQE 9

RESULT 5
US-08-159-339A-762
; Sequence 762, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 762:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-159-339A-762

Query Match 35.7%; Score 5; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFOQE 11
|||||

DB 2 GFOQE 6

RESULT 6
US-08-564-972-46
; Sequence 46, Application US/08564972
; Patent No. 5843462
; GENERAL INFORMATION:
; APPLICANT: Conti-Pine, B. M.
; TITLE OF INVENTION: DIPHTHERIA TOXIM EPITOPES
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/564,972
; FILING DATE: 30-NOV-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 600.344US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-564-972-46

Query Match 35.7%; Score 5; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFOQE 11
|||||
DB 9 GFOQE 13

RESULT 7
US-09-219-849-13
; Sequence 13, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.

; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; PREPARATION THEREOF
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Formula
; NAME/KEY: MOD_RES
; LOCATION: (3)
; OTHER INFORMATION: His or Met
; NAME/KEY: MOD_RES
; LOCATION: (6)
; OTHER INFORMATION: His or Met
; US-09-219-849-13

Query Match 35.7%; Score 5; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGFQG 10
Db 18 PGFQG 22

RESULT 8
US-08-176-500-36
; Sequence 36, Application US/08176500
; Patent No. 5498538
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents/
; NUMBER OF SEQUENCES: 141
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,500
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/013,416
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-143
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-176-500-36

Query Match 35.7%; Score 5; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DPGFQ 9
Db 18 DPGFQ 22

RESULT 9
US-08-471-052A-36
; Sequence 36, Application US/08471052A
; Patent No. 5625033
; GENERAL INFORMATION:
; APPLICANT: Kay, B. K.
; APPLICANT: Fowlkes, D. M.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 166
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,052A
; FILING DATE: 06-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-471-052A-36

Query Match 35.7%; Score 5; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DPGFQ 9
Db 18 DPGFQ 22

RESULT 10
US-08-189-331-36
; Sequence 36, Application US/08189331
; Patent No. 5747334
; GENERAL INFORMATION:

APPLICANT: Kay, B. K.
APPLICANT: Fowkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,331
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-189-331-36

Query Match 35.7%; Score 5; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0;

QY 5 DPGFQ 9
| | | | |
Db 18 DPGFQ 22

RESULT 11
US-08-471-939-36
Sequence 36, Application US/08471939
Patent No. 5844076
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,939
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/013,416
FILING DATE: 01-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-939-36

Query Match 35.7%; Score 5; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0;

QY 5 DPGFQ 9
| | | | |
Db 18 DPGFQ 22

RESULT 12
US-08-471-800-36
Sequence 36, Application US/08471800
Patent No. 5852167
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,800
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-800-36

us-08-957-709-81.rai

Thu Jan 31 13:33:01 2002

```

Query Match      35.7%; Score 5; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGFQ 9
DB 18 DPGFQ 22

RESULT 13
US-08-471-068-36
; Sequence 36, Application US/08471068
; Patent No. 5948635
; GENERAL INFORMATION:
; APPLICANT: KAY, B. K.
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents
; NUMBER OF SEQUENCES: 186
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,068
; FILING DATE: 08-SEP-1994
; PRIOR APPLICATION NUMBER: 08/189,331
; FILING DATE: 08-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-155
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-471-068-36

Query Match      35.7%; Score 5; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGFQ 9
DB 18 DPGFQ 22

RESULT 14
US-08-464-531-86
; Sequence 86, Application US/08464531
; Patent No. 5783144
; GENERAL INFORMATION:
; APPLICANT: FOWLES, Dana M.
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES:
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROADCH, Jim
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELEPHONE:
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH:
; TYPE:
; STRANDEDNESS:
; TOPOLOGY:
; MOLECULE TYPE:
; US-08-464-531-86

Query Match      35.7%; Score 5; DB 1; Length 67;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GELKL 14
DB 41 GELKL 45

RESULT 15
US-08-461-598-86
; Sequence 86, Application US/08461598
; Patent No. 5876951
; GENERAL INFORMATION:
; APPLICANT: FOWLES, Dana M.
; TITLE OF INVENTION:
; NUMBER OF SEQUENCES:
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MANFREDI, John
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELEPHONE:
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO:
; SEQUENCE CHARACTERISTICS:
; LENGTH:
; TYPE:
; STRANDEDNESS:
; TOPOLOGY:
; MOLECULE TYPE:
; US-08-461-598-86
```

APPLICANT: PAUL, Jeremy
APPLICANT: TRUEHEART, Joshua
TITLE OF INVENTION: YEAST CELLS ENGINEERED TO PRODUCE
TITLE OF INVENTION: PHEROMONE SYSTEM PROTEIN SURROGATES, AND USES THEREFOR
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,598
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/322,137
FILING DATE: 13-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/309,313
FILING DATE: 20-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/190,328
FILING DATE: 31-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/041,431
FILING DATE: 31-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: FOLWKES-2F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 86:
SEQUENCE CHARACTERISTICS:
LENGTH: 67 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-461-598-86

Query Match 35.7%; Score 5; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GELKL 14
Db 41 GELKL 45

Search completed: January 31, 2002, 13:15:16
Job time: 97 sec

This Page Blank (uspto)

GenCore version.4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:20 ; Search time 78.64 seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-81
Perfect score: 14
Sequence: 1 TGLIDPGFQGEKLK 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	287	1 WZBEP1	dUTP pyrophosphatase
2	11	78.6	287	2 T42968	hypothetical prote
3	7	50.0	289	1 S55649	deoxyuridine triph
4	6	42.9	24	2 T50123	peroxisomal target
5	6	42.9	177	1 NRASG	ribonuclease alpha
6	6	42.9	186	2 E86004	hypothetical prote
7	6	42.9	186	2 H65134	hypothetical 21.2
8	6	42.9	250	2 S64849	hypothetical prote
9	6	42.9	272	2 A72568	hypothetical prote
10	6	42.9	294	1 S73832	MG237 homolog F10
11	6	42.9	298	2 A82849	dihydropterolate sy
12	6	42.9	299	1 G64060	conserved hypothet
13	6	42.9	303	2 A84831	probable protein k
14	6	42.9	319	2 T42423	PAS7 protein homol
15	6	42.9	346	2 A58583	testosterone-resis
16	6	42.9	356	2 S16907	collagen alpha 1(I
17	6	42.9	367	2 E75031	hydrogenase expres
18	6	42.9	445	2 A96599	protein F20N2.10 l
19	6	42.9	511	2 B83148	apolipoprotein N-a
20	6	42.9	635	2 A57131	collagen alpha 2(V
21	6	42.9	669	2 D82739	exonuclease ABC s
22	6	42.9	1041	2 B81281	probable secreted
23	6	42.9	1669	1 CGHU4B	collagen alpha 1(I
24	6	42.9	1669	1 CGMS4B	collagen alpha 3(I
25	6	42.9	1670	1 CGHU3B	collagen alpha 4 c
26	6	42.9	1707	2 A33526	collagen alpha-4 c
27	6	42.9	1747	2 A54121	dipeptidyl-peptida
28	5	35.7	33	2 S77568	cell division cycl
29	5	35.7	48	2 PNO486	

30	5	35.7	59	2 B83795	hypothetical prote
31	5	35.7	65	2 S35170	cytochrome P450 (c
32	5	35.7	72	2 T11975	hypothetical prote
33	5	35.7	90	1 DNZRHU	DNA-binding protei
34	5	35.7	91	2 S37190	hypothetical prote
35	5	35.7	95	2 G64543	hypothetical prote
36	5	35.7	102	2 F75074	hypothetical prote
37	5	35.7	102	2 H71090	hypothetical prote
38	5	35.7	103	2 F75334	probable dioxygena
39	5	35.7	111	2 T12857	hypothetical prote
40	5	35.7	112	2 A69268	hypothetical prote
41	5	35.7	113	2 PC4289	brain and muscle A
42	5	35.7	118	2 T16510	hypothetical prote
43	5	35.7	135	2 E86759	hypothetical prote
44	5	35.7	141	2 A64096	universal stress p
45	5	35.7	141	2 D82368	universal stress p

ALIGNMENTS

RESULT 1
WZBEP1
dUTP pyrophosphatase (EC 3.6.1.23) - saimirine herpesvirus 1 (strain 11)
C:Species: saimirine herpesvirus 1
A:Note: host Saimiri sciureus (common squirrel monkey)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 18-Jun-1999
C:Accession: G36811
R:Albrecht, J.
submitted to the EMBL Data Library, January 1992
A:Description: Primary structure of the herpesvirus saimir genome.
A:Reference number: A36806
A:Accession: G36811
A:Molecule type: DNA
A:Residues: 1-287 <ALB>
A:Cross-references: GB:X64346; NID:g60320; PIDN:CAA45677.1; PID:g60375
R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.;
J. Virol. 66, 5047-5058, 1992
A:Title: Primary structure of the herpesvirus saimir genome.
A:Reference number: A37309; MUID:92333688
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Genetics:
A:Gene: 54
C:Superfamily: herpesvirus dUTP pyrophosphatase
C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 1; Length 287;
Best Local Similarity 100.0%; Pred. No. 6e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGLIDPGFQGEKLK 14
|||||
Db 74 TGLIDPGFQGEKLK 87

RESULT 2
T42968
hypothetical protein - ateline herpesvirus 3 (strain 73)
C:Species: ateline herpesvirus 3
A:Variety: strain 73
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
R:Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A:Description: Primary structure of the herpesvirus ateles genome.
A:Reference number: Z2274
A:Accession: T42968
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-287 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:NAC95579.1

A;Experimental source: strain 73

C;Superfamily: herpesvirus dntp pyrophosphatase

Query Match 78.6%; Score 11; DB 2; Length 287;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDPGFQGE 11
| | | | | | | | | |
DB 74 TGLIDPGFQGE 84

RESULT 3

S55649 deoxyuridine triphosphatase 54 - equine herpesvirus 2

C;Species: equine herpesvirus 2

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C;Accession: S55649

R;Teilford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.

J. Mol. Biol. 249, 520-528, 1995

A;Title: The DNA sequence of equine herpesvirus 2.

A;Reference number: S55594; MUID:95302501

A;Accession: S55649

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-289 <TEL>

A;Cross-references: GB:U20824; NID:9695172; PIDN:AAC13842.1; PID:9695227

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

C;Superfamily: herpesvirus dntp pyrophosphatase

Query Match 50.0%; Score 7; DB 1; Length 289;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDPG 7
| | | | | | | | | |
DB 77 TGLIDPG 83

RESULT 4

T50123 peroxisomal targeting signal 2 receptor (pts2 receptor) peroxin-7 [imported] - fission y

C;Species: Schizosaccharomyces pombe

C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 20-Jun-2000

C;Accession: T50123

R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, February 2000

A;Reference number: Z25039

A;Accession: T50123

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-24 <SEE>

A;Cross-references: EMBL:AL157734; PIDN:CAB75780.1; GSPDB:GN000066; SPDB:SPAC1834.12

A;Experimental source: strain 972h(-); cosmid c1834

A;Gene: SPDB:SPAC1834.12

A;Map position: 1

C;Keywords: peroxisome

Query Match 42.9%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGFQGE 11
| | | | | | | | | |
DB 7 PGFQGE 12

RESULT 5

NRAS5G

Query Match 42.9%; Score 6; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ribonuclease alpha-sarcin (EC 3.1.1.-) precursor - Aspergillus giganteus
C;Species: Aspergillus giganteus
C;Date: 18-Apr-1984 #sequence_revision 30-Sep-1991 #text_change 16-Jun-2000
C;Accession: J00138; S12582; A00802; S21866
R;Wendt, S.; Felske-Zech, H.; Henze, P.P.C.; Ulbrich, N.; Stahl, U.
Gene 124, 239-244, 1993
A;Title: Characterization of the gene encoding alpha-sarcin, a ribosome-inactivating
A;Reference number: J00138; MUID:93185929

A;Accession: J00138

A;Molecule type: DNA

A;Residues: 1-177 <WNE>

A;Cross-references: EMBL:X60770; NID:g2310; PIDN:CAA43180.1; PID:g2311

R;Oka, T.; Natori, Y.; Tsuruga, K.; Endo, Y.

Nucleic Acids Res. 18, 1897, 1990

A;Title: Complete nucleotide sequence of cDNA for the cytotoxin alpha sarcin.

A;Reference number: S12582; MUID:90245591

A;Accession: S12582

A;Molecule type: mRNA

A;Residues: 1-177 <OKA>

A;Cross-references: GB:D13704; GB:D00516; NID:g217810; PIDN:BAA02863.1; PID:g217811

R;Sacco, G.; Drickamer, K.; Wool, I.G.

J. Biol. Chem. 258, 5811-5818, 1983

A;Title: The primary structure of the cytotoxin alpha-sarcin.

A;Reference number: A00802; MUID:93213554

A;Accession: A00802

A;Molecule type: protein

A;Residues: 28-177 <SAC>

C;Comment: Alpha-sarcin is specific for purines in both single- and double-stranded R the 60S subunit of ribosomes.

C;Genetics:

A;Gene: sar

A;Introns: 29/3

C;Superfamily: ribonuclease T1

C;Keywords: endonuclease; extracellular protein; hydrolase; toxin

F;1-27/Domain: signal sequence #status predicted <SIG>

F;28-177/Product: ribonuclease alpha-sarcin #status experimental <MAT>

F;33-175,103-159/Disulfide bonds: #status predicted

F;75,77,123,148,164/Active site: Tyr, His, Glu, Arg, His #status predicted

Query Match 42.9%; Score 6; DB 1; Length 177;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QGELKL 14
| | | | | | | | | |
DB 169 QGELKL 174

RESULT 6

E86004 hypothetical protein yrfE [imported] - Escherichia coli (strain O157:H7)

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C;Accession: E86004

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apoda

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: E86004

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-186 <STO>

A;Cross-references: GB:AE005174; NID:g12518020; PIDN:AAG58497.1; GSPDB:GN00145; UWGP:

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: yrfE

QY 2 GLIDPG 7
|||||
Db 80 GLIDPG 85

RESULT 7
H65134

hypothetical 21.2 kD protein in mrca-pckA intergenic region - Escherichia coli (strain K12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: H65134
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
A:Residues: 1-186 <BLAT>
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A6720; MUID:97426617
A:Accession: H65134
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-186 <BLAT>
A:Cross-references: GB:AE000415; GB:U00096; NID:gl789798; PIDN:AACT6422.1; PID:gl789800;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yrfE

Query Match 42.9%; Score 6; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDPG 7
|||||
Db 80 GLIDPG 85

RESULT 8
S64849

hypothetical protein YLR022c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L1722
C:Species: Saccharomyces cerevisiae
C:Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 29-Oct-1999
R:Obermaier, B.; Piravandi, E.; Rinke, M.
A:Reference number: S64845
A:Accession: S64849
A:Molecule type: DNA
A:Residues: 1-250 <OBEL>
A:Cross-references: EMBL:Z73194; NID:gl360327; PIDN:CAA97545.1; PID:e245773; PID:gl360327
A:Experimental source: strain S288C
R:Vandenbol, M.; Portetelle, D.; Hilger, F.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64742
A:Accession: S64844
A:Molecule type: DNA
A:Residues: 135-250 <VAN>
A:Cross-references: EMBL:Z73194; MIPS:YLR022c
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 12R

Query Match 42.9%; Score 6; DB 2; Length 250;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDP 6
|||||
Db 215 TGLIDP 220

RESULT 9

A72568
hypothetical protein APE1827 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72568
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Tawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: A72568
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-272 <RAW>
A:Cross-references: DDBJ:AP000062; NID:gs105244; PIDN:BAA80830.1; PID:d1044616; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1827

Query Match 42.9%; Score 6; DB 2; Length 272;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFQGEI 12
|||||
Db 231 GFQGEI 236

RESULT 10
S73832

MG237 homolog F10_orf294 - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 10-Dec-1999
C:Accession: S73832
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885
A:Accession: S73832
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-294 <HIM>
A:Cross-references: EMBL:AE000050; GB:U00089; NID:gl674157; PIDN:AAB96154.1; PID:gl67
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: Mycoplasma hypothetical protein MG237

Query Match 42.9%; Score 6; DB 1; Length 294;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGFQGE 11
|||||
Db 170 PGFQGE 175

RESULT 11
A82849

dihydropteroate synthase XF0091 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: A82849
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: A82849
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-298 <SIM>
A:Cross-references: GB:AE003863; GB:AE003849; NID:g9104863; PIDN:AAF82904.1; GSPDB:GN00139
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. et al.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig, Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B. et al.; Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak, A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tsuchioka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0091
C:Superfamily: dihydropteroate synthase; dihydropteroate synthase homology

Query Match 42.9%; Score 6; DB 2; Length 298;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 LIDPGF 8
Db 187 LIDPGF 192

RESULT 12
G64060
conserved hypothetical protein H10301 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: G64060
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A. et al.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. et al.; Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: G64060
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-299 <TIGR>
A:Cross-references: GB:U32716; GB:L42023; NID:gl573268; PIDN:AAC21566.1; PID:gl573270; T
A:Experimental source: strain Rd KW20
C:Superfamily: hypothetical protein b0558

Query Match 42.9%; Score 6; DB 1; Length 299;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 FQELK 13
Db 27 FQELK 32

RESULT 13
A84831
probable protein kinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84831
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. et al.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999

A:Molecule type: DNA
A:Residues: 1-303 <STO>
A:Cross-references: GB:AE002093; NID:g2651301; PIDN:AAB87581.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g40560
A:Map position: 2
C:Superfamily: Aquifex aeolicus serine/threonine-specific protein kinase; protein kin

Query Match 42.9%; Score 6; DB 2; Length 303;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GLIDPG 7
Db 199 GLIDPG 204

RESULT 14
T42423
PAS7 protein homolog - fission yeast (Schizosaccharomyces pombe) (fragment)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42423
R:Toshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A:Reference number: Z17323; MUID:98162722
A:Accession: T42423
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-319 <YOS>
A:Cross-references: EMBL:D89147; NID:gl749501; PIDN:BAAL3809.1; PID:gl749502
A:Experimental source: strain PR745

Query Match 42.9%; Score 6; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 PGFOGE 11
Db 18 PGFOGE 23

RESULT 15
A58583
testosterone-resistant immunity-associated protein IAP38 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 25-Apr-1997 #sequence_revision 09-May-1997 #text_change 08-Oct-1999
C:Accession: A58583
R:Kruecken, J.; Schmitt-Wrede, H.P.; Markmann-Mullisch, U.; Wunderlich, F.
Biochem. Biophys. Res. Commun. 230, 167-170, 1997
A:Title: Novel gene expressed in spleen cells mediating acquired testosterone-resista
A:Reference number: A58583; MUID:97148595
A:Accession: A58583
A:Molecule type: mRNA
A:Residues: 1-346 <KRU>
A:Cross-references: GB:Y08026; NID:gl550784; PIDN:CAA69283.1; PID:c266428; PID:gl5507
A:Experimental source: spleen cell
C:Comment: This protein is a plasma membrane protein with two membrane-spanning domai
chabaudi malaria.
C:Genetics:
A:Gene: iap38
F:148-167/Domain: transmembrane #status predicted <TM1>
F:320-335/Domain: transmembrane #status predicted <TM2>

Query Match 42.9%; Score 6; DB 2; Length 346;

Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGFQGE 11
 |||||
Db 21 PGFQGE 26

Search completed: January 31, 2002, 13:20:20
Job time: 121 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:52 ; Search time 130.99 Seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-81

Perfect score: 14

Sequence: 1 TGLIDPGFQGECLK 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	78.6	287	12 Q9YTL2	Q9YTL2 ateline her
2	7	50.0	289	12 Q66656	Q66656 equine herp
3	7	50.0	299	12 P88991	P88991 murid herpe
4	6	42.9	125	10 Q9LN05	Q9LN05 arabisdopsi
5	6	42.9	159	10 Q9SRR8	Q9SRR8 arabisdopsi
6	6	42.9	162	12 Q65243	Q65243 african swi
7	6	42.9	165	12 Q65199	Q65199 african swi
8	6	42.9	177	3 P87063	P87063 aspergillus
9	6	42.9	245	2 Q9KY25	Q9KY25 streptomyc
10	6	42.9	246	11 P70224	P70224 mus musculu
11	6	42.9	272	1 Q9YAW8	Q9YAW8 aeropyrum p
12	6	42.9	288	4 Q9BZ45	Q9BZ45 homo sapien
13	6	42.9	298	2 Q9PH55	Q9PH55 xylella fas
14	6	42.9	303	10 Q22879	Q22879 arabisdopsi
15	6	42.9	321	6 Q9BE81	Q9BE81 macaca fasc
16	6	42.9	333	2 Q9LBV7	Q9LBV7 synechococc
17	6	42.9	337	2 Q9XBD3	Q9XBD3 amycolatops
18	6	42.9	338	10 Q9SQ06	Q9SQ06 arabisdopsi
19	6	42.9	367	1 Q9UZB3	Q9UZB3 pyrococcus

20	6	42.9	377	13 Q9IAD4	Q9IAD4 xenomystus
21	6	42.9	401	2 Q69141	Q69141 streptococc
22	6	42.9	401	2 Q99XV1	Q99XV1 streptococc
23	6	42.9	418	10 Q9LTT4	Q9LTT4 arabisdopsi
24	6	42.9	428	2 Q9FB00	Q9FB00 streptococc
25	6	42.9	445	10 Q9LG00	Q9LG00 arabisdopsi
26	6	42.9	445	10 Q9C5N4	Q9C5N4 arabisdopsi
27	6	42.9	456	10 Q9FKS9	Q9FKS9 arabisdopsi
28	6	42.9	474	10 Q9LWJ1	Q9LWJ1 oryza sativ
29	6	42.9	568	2 Q9CP14	Q9CP14 pasteurella
30	6	42.9	669	2 Q9PER1	Q9PER1 xylella fas
31	6	42.9	673	2 Q9RBK2	Q9RBK2 xanthomonas
32	6	42.9	749	2 Q9A462	Q9A462 caulobacter
33	6	42.9	855	3 Q9P8D5	Q9P8D5 penicillium
34	6	42.9	979	13 Q9I9K3	Q9I9K3 gallus gall
35	6	42.9	1041	2 Q9PMT5	Q9PMT5 campylobact
36	6	42.9	1048	5 Q9VX90	Q9VX90 drosophila
37	6	42.9	1147	5 Q9GUT9	Q9GUT9 leishmania
38	6	42.9	1362	4 Q9ULF8	Q9ULF8 homo sapien
39	6	42.9	1478	5 Q9V6W9	Q9V6W9 drosophila
40	6	42.9	1621	4 Q9H4R9	Q9H4R9 homo sapien
41	6	42.9	1670	4 Q9BQT2	Q9BQT2 homo sapien
42	6	42.9	1747	5 Q26640	Q26640 strongyloce
43	5	35.7	38	2 P96300	P96300 alcaligenes
44	5	35.7	54	4 Q9UML5	Q9UML5 homo sapien
45	5	35.7	59	2 Q9KDP9	Q9KDP9 bacillus ha

ALIGNMENTS

RESULT 1
Q9YTL2 PRELIMINARY; PRT; 287 AA.
ID Q9YTL2
AC Q9YTL2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF 54.
OS Ateeline herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=85618;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=73;
RA Albrecht J.-C., Fleckenstein B.;
RT "Primary Structure of the Herpesvirus Ateles Genome."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF083424; AAC95579.1; .
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
SQ SEQUENCE 287 AA; 32154 MW; 3A819076115FC280 CRC64;

Query Match 78.6%; Score 11; DB 12; Length 287;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDPGFQGE 11
| | | | | | | | | |
Db 74 TGLIDPGFQGE 84

RESULT 2
Q66656 PRELIMINARY; PRT; 289 AA.
ID Q66656
AC Q66656;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF 54.
OS Equine herpesvirus type 2 (strain 86/87).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=82831;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95302501; PubMed=7783207;
RA Telford E.A., Watson M.S., Aird H.C., Perry J., Davison A.J.;
RT "The DNA sequence of equine herpesvirus 2.";
RL J. Mol. Biol. 249:520-528(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Telford E.A.R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20824; AAC13842.1; -;
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
SQ SEQUENCE 289 AA; 32957 MW; 6FCEB405630FA996 CRC64;

Query Match 50.0%; Score 7; DB 12; Length 289;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDPG 7
Db 77 TGLIDPG 83

RESULT 3
ID P88991 PRELIMINARY; PRT; 299 AA.
AC P88991; 1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DSOXYURIDINE TRIPHOSPHATASE.
GN BORFI OR GAMMAHV. ORF54.
OS murid herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=33708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-G24;
RA Mackett M., Stewart J.P., Pepper V., Chee M., Efstathiou S.,
RA Nash A.A., Arrand J.R.;
RL J. Gen. Virol. 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-WUMS, AND G2.4;
RX MEDLINE=97366649; PubMed=9223479;
RA Virgin H.W. IV, Latreille P., Wamsley P., Hallsworth K., Weck K.E.,
RA Dal Canto A.J., Speck S.H.;
RP "Complete sequence and genomic analysis of murine gammaherpesvirus
68.";
RL J. Virol. 71:5894-5904(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-WUMS;
RA Latreille P., Wamsley P., Waterston R.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-G2.4;
RX MEDLINE=97335274; PubMed=9191940;
RA Mackett M., Stewart J.P., de V Pepper S., Chee M., Efstathiou S.,
RA Nash A.A., Arrand J.R.;
RP "Genetic content and preliminary transcriptional analysis of a
representative region of murine gammaherpesvirus 68.";
RL J. Gen. Virol. 78:1425-1433(1997).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-G2.4;

RA Milligan S., Efstathiou S., Stewart J.P., Nash A.A., Davison A.J.;
RT "Genetic content of murine gammaherpesviruses";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y09060; CAA70249.1; -;
DR EMBL; U37553; AAB66412.1; -;
DR EMBL; AF105037; AAL19318.1; -;
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
SQ SEQUENCE 299 AA; 32657 MW; 7B996E5D1B77FB3B CRC64;

Query Match 50.0%; Score 7; DB 12; Length 299;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDPG 7
Db 81 TGLIDPG 87

RESULT 4
ID Q9LN05 PRELIMINARY; PRT; 125 AA.
AC Q9LN05;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE T6D22.11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC Eurosidia II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T6D22 from chromosome
I.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Lam B.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC026875; AAF79825.1; -;
SQ SEQUENCE 125 AA; 13865 MW; EE7040BD198C3935 CRC64;

Query Match 42.9%; Score 6; DB 10; Length 125;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDPGFQ 9
Db 11111

Db 116 IDPGFQ 121

RESULT 5
 Q9SRR8 PRELIMINARY; PRT; 159 AA.
 AC Q9SRR8;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE F2103.19 PROTEIN.
 GN F2103.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III BAC F2103 genomic sequence.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC009853; AAF02167.1;
 DR HSP; P00259; IGXP.
 SQ SEQUENCE 159 AA; 17602 MW; AFDA3BE19E011EE0 CRC64;

Query Match 42.9%; Score 6; DB 10; Length 159;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDP 6
 DB 75 TGLIDP 80

RESULT 6
 Q65243 PRELIMINARY; PRT; 162 AA.
 AC Q65243;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE DEOXYRIDINE TRIPHOSPHATE.
 OS African swine fever virus (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
 OC African swine fever-like viruses.
 OX NCBI_TaxID=10497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MALAWI LIL20 /1;
 RX MEDLINE=94014996; PubMed=8409937;
 RA Ydelingum S., Baylis S.A., Bristow C., Smith G.L., Dixon L.K.;
 RT "Duplicated genes within the variable right end of the genome of a
 pathogenic isolate of African swine fever virus.";
 RL J. Gen. Virol. 74:2125-2130(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MALAWI LIL20 /1;
 RX MEDLINE=94292916; PubMed=8021596;
 RA Dixon L.K., Twigg S.R.F., Baylis S.A., Ydelingum S., Bristow C.,
 RA Hammond J.M., Smith G.L.;
 RT "Nucleotide sequence of a 55 kbp region from the right end of the
 genome of a pathogenic African swine fever virus isolate (Malawi
 LIL20/1).";
 RL J. Gen. Virol. 7:1655-1684(1994).
 DR EMBL; X71982; CA450839.1;
 DR InterPro; IPR001428; dUTPase.
 DR Pfam; PF00692; dUTPase; 1.
 SQ SEQUENCE 162 AA; 17872 MW; FFD71331454A5737 CRC64;

Query Match 42.9%; Score 6; DB 12; Length 162;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDPG 7
 DB 88 GLIDPG 93

RESULT 7
 Q65199 PRELIMINARY; PRT; 165 AA.
 ID Q65199;
 AC Q65199;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE DUTPASE.
 GN E165R.
 OS African swine fever virus (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae;
 OC African swine fever-like viruses.
 OX NCBI_TaxID=10497;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=96036500; PubMed=7483270;
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C.,
 RA Rodriguez J.F., Vinuela E.;
 RT "Immune protection conferred by the baculovirus-related glycoprotein
 of Thogoto virus (Orthomyxoviridae).";
 RL Virology 208:249-278(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=94233765; PubMed=8178480;
 RA La Vega I., Gonzalez A., Blasco R., Calvo V., Vinuela E.;
 RT "Nucleotide sequence and variability of the inverted terminal
 repetitions of African swine fever virus DNA.";
 RL Virology 201:152-156(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=90219205; PubMed=2325203;
 RA Gonzalez A., Calvo V., Almazan F., Almendral J.M., Ramirez J.C.,
 RA La Vega I., Blasco R., Vinuela E.;
 RT "Multigene families in African swine fever virus: family 360.";
 RL J. Virol. 64:2073-2081(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=90219204; PubMed=2325202;
 RA Almendral J.M., Almazan F., Blasco R., Vinuela E.;
 RT "Multigene families in African swine fever virus: family 110.";
 RL J. Virol. 64:2064-2072(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=91134988; PubMed=1994575;
 RA Camacho A., Vinuela E.;
 RT "protein p22 of African swine fever virus: an early structural protein
 that is incorporated into the membrane of infected cells.";
 RL Virology 181:251-257(1991).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=91134988; PubMed=1994575;
 RA Almazan F., Murguía J.R., Rodriguez J.M., La Vega I., Vinuela E.;
 RL J. Gen. Virol. 0:0-0(0).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BA71V;
 RX MEDLINE=94187118; PubMed=8139051;

RA Rodriguez J.M., Yanez R.J., Pan R., Rodriguez J.F., Salas M.L.,
RA Vinuela E.;
RA "Multigene families in African swine fever virus: family 505.";
RT J. Virol. 68:2746-2751(1994).
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93346971; PubMed=8393914;
RA Yanez R.J., Rodriguez J.M., Rodriguez J.F., Salas M.L., Vinuela E.;
RT "African swine fever virus thymidylate kinase gene: sequence and
transcriptional mapping.";
RL J. Gen. Virol. 74:1633-1638(1993).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=94065656; PubMed=8245848;
RA Alcamí A., Angulo A., Vinuela E.;
RT "Mapping and sequence of the gene encoding the African swine fever
viral protein of M(r) 11500.";
RL J. Gen. Virol. 74:2317-2324(1993).
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93277388; PubMed=8503790;
RA Munoz M., Freije J.M., Salas M.L., Vinuela E., Lopez-Otin C.;
RT "Structure and expression in E. coli of the gene coding for protein
p10 of African swine fever virus.";
RL Arch. Virol. 130:93-107(1993).
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=90357780; PubMed=2389555;
RA Blasco R., Lopez-Otin C., Munoz M., Bockamp E.O., Simon-Mateo C.,
RA Vinuela E.;
RT "Sequence and evolutionary relationships of African swine fever virus
thymidine kinase.";
RL Virology 178:301-304(1990).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93281390; PubMed=8506138;
RA Yanez R.J., Boursnell M., Nogal M.L., Yuste L., Vinuela E.;
RT "African swine fever virus encodes two genes which share significant
homology with the two largest subunits of DNA-dependent RNA
polymerases.";
RL Nucleic Acids Res. 21:2423-2427(1993).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93353606; PubMed=8102411;
RA Rodriguez J.M., Yanez R.J., Almazan F., Vinuela E., Rodriguez J.F.;
RT "African swine fever virus encodes a CD2 homolog responsible for the
adhesion of erythrocytes to infected cells.";
RL J. Virol. 67:5312-5320(1993).
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=94085774; PubMed=8262374;
RA Yanez R.J., Rodriguez J.M., Boursnell M., Rodriguez J.F., Vinuela E.;
RT "Two putative African swine fever virus helicases similar to yeast
'DEAH' pre-mRNA processing proteins and vaccinia virus ATPases D1LL
and D6R.";
RL Gene 134:161-174(1993).
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=90223993; PubMed=2327074;
RA Lopez-Otin C., Freije J.M., Parra F., Mendez E., Vinuela E.;
RT "Mapping and sequence of the gene coding for protein p72, the major
capsid protein of African swine fever virus.";
RL Virology 175:477-484(1990).
RN [16]
RP SEQUENCE FROM N.A.

RC STRAIN=BA71V;
RX MEDLINE=94123986; PubMed=8293992;
RA Rodriguez J.M., Yanez R.J., Rodriguez J.F., Vinuela E., Salas M.L.;
RT "The DNA polymerase-encoding gene of African swine fever virus:
sequence and transcriptional analysis.";
RL Gene 136:103-110(1993).
RN [17]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93327788; PubMed=8335009;
RA Simon-Mateo C., Andres G., Vinuela E.;
RT "Polyprotein processing in African swine fever virus: a novel gene
expression strategy for a DNA virus.";
RL EMBO J. 12:2977-2987(1993).
RN [18]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93233210; PubMed=8474154;
RA Prados F.J., Vinuela E., Alcamí A.;
RT "Sequence and characterization of the major early phosphoprotein p32
of African swine fever virus.";
RL J. Virol. 67:2475-2485(1993).
RN [19]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=92260660; PubMed=1583732;
RA Alcamí A., Angulo A., Lopez-Otin C., Munoz M., Freije J.M.,
RA Carrascosa A.L., Vinuela E.;
RT "Amino acid sequence and structural properties of protein p12, an
African swine fever virus attachment protein.";
RL J. Virol. 66:3860-3868(1992).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174976; PubMed=8438592;
RA Yanez R.J., Vinuela E.;
RT "African swine fever virus encodes a DNA ligase.";
RL Virology 193:531-536(1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=93174941; PubMed=8382399;
RA Pena L., Yanez R.J., Revilla Y., Vinuela E., Salas M.L.;
RT "African swine fever virus guanylyltransferase.";
RL Virology 193:319-328(1993).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=95159428; PubMed=7856088;
RA Simon-Mateo C., Freije J.M., Lopez-Otin C., Blasco R., Vinuela E.;
RT "Mapping and sequence of the gene encoding protein p17, a major
African swine fever virus structural protein.";
RL Virology 206:1140-1144(1995).
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=92263807; PubMed=1316688;
RA Garcia-Beato R., Freije J.M., Lopez-Otin C., Blasco R., Vinuela E.,
RA Salas M.L.;
RT "A gene homologous to topoisomerase II in African swine fever virus.";
RL Virology 188:938-947(1992).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN=BA71V;
RX MEDLINE=94091056; PubMed=8266720;
RA Freije J.M., Lain S., Vinuela E., Lopez-Otin C.;
RT "Nucleotide sequence of a nucleoside triphosphate phosphohydrolase
Query Match 42.9%; Score 6; DB 12; Length 165;
Best Local Similarity 100.0%; Pred No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 88 GLIDPG 93
|||||

RESULT 8
P87063 PRELIMINARY; PRT; 177 AA.
AC P87063;
DT 01-JUL-1997 (TREMELREL. 04, Created)
DT 01-JUL-1997 (TREMELREL. 04, Last sequence update)
DT 01-JUN-2000 (TREMELREL. 14, Last annotation update)
DE GIGANTIN PRECURSOR.
GN GIG.
OS Aspergillus giganteus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5060;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 5818;
RX MEDLINE=97367962; PubMed=9224729;
RA Wirth J., Martinez del Pozo A., Mancheno J.M., Martinez-Ruiz A.,
RA Lacadena J., Onaderra M., Gavilanes J.G.;
RT "Sequence determination and molecular characterization of giantin, a
RT cytotoxic protein produced by the mould Aspergillus giganteus IFO
RT 5818.";
RL Arch. Biochem. Biophys. 343:188-193(1997).
DR EMBL; Y08457; CAA69713.1; -;
DR HSSP; P04389; IAQZ.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 177 POTENTIAL.
SQ SEQUENCE 177 AA; 19827 MW; 231EE906F08E46A0 CRC64;

Query Match 42.9%; Score 6; DB 3; Length 177;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 QGELKL 14
|||||
Db 169 QGELKL 174

RESULT 9
Q9KY25 PRELIMINARY; PRT; 245 AA.
AC Q9KY25;
DT 01-OCT-2000 (TREMELREL. 15, Created)
DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)
DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)
DE HYPOTHETICAL 27.4 KDA PROTEIN.
GN SCC8A.23C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL358892; CAB92841.1; -;
KW Hypothetical protein.
SQ SEQUENCE 245 AA; 27432 MW; EAE5956CD22B69DE CRC64;

Query Match 42.9%; Score 6; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GFQEL 12
|||||
Db 220 GFQEL 225

RESULT 10
P70224 PRELIMINARY; PRT; 246 AA.
AC P70224;
DT 01-FEB-1997 (TREMELREL. 02, Created)
DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)
DT 01-OCT-2000 (TREMELREL. 15, Last annotation update)
DE IMMUNE ASSOCIATED PROTEIN 38.
GN IAP38 OR IMAP38.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10; TISSUE=SPLEEN;
RX MEDLINE=97148595; PubMed=9020038;
RA Kruecken J., Schmitt-Wrede H.P., Markmann-Mulisch U., Wunderlich F.;
RT "Novel gene expressed in spleen cells mediating acquired testosterone-
RT resistant immunity to Plasmodium chabaudi malaria.";
RL Biochem. Biophys. Res. Commun. 230:167-170(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10; TISSUE=SPLEEN;
RX Kruecken J., Schmitt-Wrede H.P., Markmann-Mulisch U., Wunderlich F.;
RT "Novel gene expressed in spleen cells mediating acquired testosterone-
RT resistant immunity to Plasmodium chabaudi malaria.";
RL Biochem. Biophys. Res. Commun. 230:167-170(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10; TISSUE=SPLEEN;
RX Kruecken J.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10; TISSUE=SPLEEN;
RX Kruecken J.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10, AND 129/OLA; TISSUE=SPLEEN;
RX MEDLINE=99377081; PubMed=10446218;
RA Kruecken J., Stamm O., Schmitt-Wrede H.P., Mincheva A., Lichter P.,
RA Wunderlich F.;
RT "Spleen-specific Expression of the Malaria-Inducible Intronless Mouse
RT Gene inap38.";
RL J. Biol. Chem. 274:24383-24391(1999).
RN [5]

Query Match 42.9%; Score 6; DB 11; Length 246;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGFQGE 11

```

Db      21  PGFGE 26
|||||
RESULT 11
Q9YAW8 PRELIMINARY; PRT; 272 AA.
ID Q9YAW8
AC Q9YAW8
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 30.0 KDA PROTEIN APE1827.
GN APE1827.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
RA Hosoyma A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamaki A., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
RL EMBL: AP000662; BAA0830.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 272 AA; 29959 MW; 740562B65DF32A96 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFQGE 12
|||||
Db 231 GFQGE 236

RESULT 12
Q9B245 PRELIMINARY; PRT; 288 AA.
ID Q9B245
AC Q9B245
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE BA307L3.1 (SIMILAR TO NUCLEAR PROTEIN NP95) (FRAGMENT).
GN BA307L3.1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Laird G.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: ALI33480; CAC22601.1; -.
KW Nuclear protein.
FT NON_TER 288
SQ SEQUENCE 288 AA; 32466 MW; C0C04FD577172033 CRC64;

Query Match 42.9%; Score 6; DB 4; Length 288;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDPGF 8
|||||
Db 187 LIDPGF 192

Query Match 42.9%; Score 6; DB 2; Length 298;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDPGF 8
|||||
Db 187 LIDPGF 192

RESULT 14
Q22879 PRELIMINARY; PRT; 303 AA.
ID Q22879
AC Q22879
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

```


DE PUTATIVE SERINE/THREONINE PROTEIN KINASE.
 GN T2P4.9.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Ketchum K.A., Lin X., Crosby M.L., Brandon R.C.,
 RA Sykes S.W., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Somerville C.R., Venter J.C.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: AC002336; AAB87581.1; -;
 DR Mendel; 26527; Arath; 3248; 26527.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_kin_actsite.
 DR Pfam: PF00069; pkinase; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 303 AA; 34936 MW; 1CF1723CA7462A89 CRC64;

Query Match 42.9%; Score 6; DB 10; Length 303;
 Best Local Similarity 100.0%; Pred.No.74;
 Matches 6; Conservative. 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDPG 7
 |||||
 Db 199 GLIDPG 204

RESULT 15
 Q9BE81 PRELIMINARY; PRT; 321 AA.
 AC Q9BE81
 DT 01-JUN-2001 (TREMELrel. 17, Created)
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMELrel. 17, Last annotation update)
 DE HYPOTHETICAL 36.5 KDA PROTEIN.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
 OC Cercopitheciinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN PARIETAL LOBE;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 RT libraries."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB056765; BAB39313.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 321 AA; 36464 MW; 1DED0CC0E0453E2E CRC64;

Query Match 42.9%; Score 6; DB 6; Length 321;
 Best Local Similarity 100.0%; Pred.No.78;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGFQGE 11
 |||||
 Db 26 PGFQGE 31

Search completed: January 31, 2002, 13:37:54
 Job time: 179 sec

GenCore version 4.5
Copyright (c) 1993 - 2000. CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:25 ; Search time 46.78 Seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-81
Perfect score: 14
Sequence: 1 TGLIDPGFGELK 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	287	1	DUT_HSVSA
2	6	42.9	177	1	RNAS_ASPGI
3	6	42.9	186	1	NUDE_ECOLI
4	6	42.9	250	1	Y222_YEAST
5	6	42.9	284	1	Y237_MYCPN
6	6	42.9	299	1	CORC_HAEIN
7	6	42.9	300	1	CORC_PASMU
8	6	42.9	308	1	PEX7_SCHPO
9	6	42.9	511	1	LNT_PSEAE
10	6	42.9	635	1	CA28_HUMAN
11	6	42.9	1214	1	SIP1_HUMAN
12	6	42.9	1669	1	CA14_HUMAN
13	6	42.9	1669	1	CA14_MOUSE
14	6	42.9	1670	1	CA34_HUMAN
15	6	42.9	1707	1	CA24_MOUSE
16	5	35.7	72	1	RPOZ_CYACA
17	5	35.7	90	1	DBH_RHIME
18	5	35.7	112	1	Y145_ARCFU
19	5	35.7	141	1	USPA_HAEIN
20	5	35.7	143	1	USPA_ECOLI
21	5	35.7	172	1	YKKB_BACSU
22	5	35.7	183	1	ATPD_CYACA
23	5	35.7	190	1	DCD_MYCTU
24	5	35.7	191	1	DCD_STRCO
25	5	35.7	191	1	PIC2_AGRTU
26	5	35.7	197	1	DCD_METTH
27	5	35.7	206	1	Y888_METJA
28	5	35.7	219	1	YAO3_SCHPO
29	5	35.7	223	1	UNG_BORBU
30	5	35.7	226	1	RNC_BUCAT
31	5	35.7	228	1	TEDT_ALCEU
32	5	35.7	241	1	YWBG_BACSU
33	5	35.7	243	1	YQEF_BACSU

34	5	35.7	246	1	NCAP_SFSV
35	5	35.7	250	1	QC97_HUMAN
36	5	35.7	253	1	PCRB_THEAC
37	5	35.7	254	1	FAEL_ECOLI
38	5	35.7	254	1	GTXC_ORYSA
39	5	35.7	255	1	UPPS_METTH
40	5	35.7	263	1	KA9_STRRI
41	5	35.7	263	1	YMFP_ECOLI
42	5	35.7	264	1	DP3A_SACER
43	5	35.7	272	1	PSBO_SINEN
44	5	35.7	273	1	NUFM_NEUCR
45	5	35.7	282	1	AQUA_ATRCA

ALIGNMENTS

RESULT 1	
DUT_HSVSA	
ID DUT_HSVSA	STANDARD; PRT; 287 AA.
AC Q01034;	
DT 01-APR-1993 (Rel. 25, Created)	
DT 01-APR-1993 (Rel. 25, Last sequence update)	
DT 15-DEC-1998 (Rel. 37, Last annotation update)	
DE DEOXYRIDINE 5'-TRIPHOSPHATE NUCLEOTIDHYDROLASE (EC 3.6.1.23)	
DE (DUTPASE) (DUTP PYROPHOSPHATASE).	
GN 54 OR EDRE3.	
OS Herpesvirus saimiri (strain 11).	
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;	
OC Gammaherpesvirinae; Rhadinovirus.	
OX NCBI_TaxID=10383;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=92333688; PubMed=1321287;	
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,	
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,	
RA Honess R.W.;	
RT "Primary structure of the herpesvirus saimiri genome.";	
RL J. Virol. 66:5047-5058(1992).	
RN [2]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=92230228; PubMed=1314457;	
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;	
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of	
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic	
RT organization between HVS and Epstein-Barr virus.";	
RL Virology 188:296-310(1992).	
CC -!- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT	
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES	
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT	
CC URACIL CANNOT BE INCORPORATED INTO DNA.	
CC -!- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.	
CC -!- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
CC EMBL; X64346; CAA45677.1; ..	
DR EMBL; M86409; AAA46131.1; ..	
DR PIR; G36811; WZBEPL.	
DR InterPro; IPR001428; dUTPase.	
DR Pfam; PF00692; dUTPase; 1.	
KW Hydrolase; Nucleotide metabolism.	
SQ SEQUENCE 287 AA; 32507 MW; 598D495D74274A11 CRC64;	

Query Match 100.0%; Score 14; DB 1; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TGLIDPGFQGLKL 14
DB 74 TGLIDPGFQGLKL 87

RESULT 2
RNAS_ASPGI
ID RNAS_ASPGI STANDARD: PRT; 177 AA.
AC P0655;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DE 01-DEC-1992 (Rel. 24, Last annotation update)
DE RIBONUCLEASE ALPHA-SARCIN PRECURSOR (EC 3.1.27.10).
GN SAR.
OS Aspergillus giganteus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5060;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245591; PubMed=2336369;
RA Oka T., Natori Y., Tanaka S., Tsurugi K., Endo Y.;
RT "Complete nucleotide sequence of cDNA for the cytotoxin alpha
RT sarcin.";
RL Nucleic Acids Res. 18:1897-1897(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=MDH 18894;
RA Wenzel S., Feiske H., Henze P., Ulbrich N., Stahl U.;
RL [3] Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE OF 28-177.
RX MEDLINE=83213554; PubMed=6343394;
RA Sacco G., Drickamer K., Wool I.G.;
RT "The primary structure of the cytotoxin alpha-sarcin.";
RL J. Biol. Chem. 258:5811-5818(1983).
CC -1- FUNCTION: ALPHA-SARCIN IS SPECIFIC FOR PURINES IN BOTH SINGLE- AND
CC DOUBLE-STRANDED RNA. ITS TOXIC ACTION ON EUKARYOTIC CELLS IS THE
CC RESULT OF CLEAVAGE OF A SINGLE PHOSPHODIESTER BOND IN THE 60S
CC SUBUNIT OF RIBOSOMES.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE PHOSPHODIESTER LINKAGE
CC BETWEEN GUANOSINE AND ADENOSINE RESIDUES AT ONE SPECIFIC POSITION
CC IN THE 28S RNA FROM RAT RIBOSOMES.
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE U2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13704; BAA02863.1; -
DR EMBL; X60770; CAA43180.1; -
DR PIR; S12582; NRASG.
DR PIR; S21866; S21866.
DR HSP; P04389; IAOZ.
KW Hydrolase; Nuclease; Protein synthesis inhibitor; Signal.
FT SIGNAL 1
FT CHAIN 28 177 RIBONUCLEASE ALPHA-SARCIN.
FT DISULFID 33 175
FT DISULFID 103 159
FT ACT_SITE 123 123
SQ SEQUENCE 177 AA; 19724 MW; 6C711D9482DC9DD1 CRC64;

```

Query Match 42.9%; Score 6; DB 1; Length 177;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QGELKL 14
 DB 169 QGELKL 174

```

RESULT 3
NUDE_ECOLI
ID NUDE_ECOLI STANDARD: PRT; 186 AA.
AC P45799;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADP COMPOUNDS HYDROLASE NUDE (EC 3.6.1.1).
GN NUDE OR B3397
OS Escherichia coli
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=98123081; PubMed=9452430;
RA O'Handley S.F., Frick D.N., Dunn C.A., Bessman M.J.;
RT "Orf186 represents a new member of the Nudix hydrolases, active on
RT adenosine(5')triphospho(5')adenosine, ADP-ribose, and NADH.";
RL J. Biol. Chem. 273:3192-3197(1998).
CC -1- FUNCTION: ACTIVE ON ADENOSINE(5')TRIPHOSPHO(5')ADENOSINE (AP3A),
CC ADP-RIBOSE, NADH, ADENOSINE(5')DIPHOSPHO(5')ADENOSINE (AP2A).
CC -1- CATALYTIC ACTIVITY: ADP-RIBOSE + H(2)O -> AMP + D-RIBOSE 5-
CC PHOSPHATE.
CC -1- COFACTOR: MAGNESIUM OR OTHER DIVALENT CATIONS.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE NUDIX HYDROLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18997; AAA58194.1; -
DR EMBL; AE000415; AAC78422.1; -
DR Ecogene; EGI2926; nude.
DR InterPro; IPR000086; NUDIX_hydrolase.
DR Pfam; PF00293; nudt; 1.
DR PRINTS; PR00502; MUTTDOMAIN.
DR PROSITE; PS00893; NUDIX; 1.
KW Hydrolase; Magnesium; Complete proteome.
FT DOMAIN 80 101 NUDIX BOX.
SQ SEQUENCE 186 AA; 21153 MW; D959AD8ECF73FCC4 CRC64;

```

Query Match 42.9%; Score 6; DB 1; Length 186;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDPG 7
 DB 80 GLIDPG 85

RESULT 4

```

YL22_YEAST
ID YL22_YEAST STANDARD; PRT; 250 AA.
AC Q07953;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 28.3 KDA PROTEIN IN PPRI-SNE7 INTERGENIC REGION.
GN YL0202C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 134-250 FROM N.A.
RA Vandenbol M., Portetelle D., Hilger F.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE UPF0023 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 273194; CAA97545.1; -
CC DR SGD; S0004012; YLR022C.
CC DR InterPro: IPR002140; UPF0023.
CC DR Pfam; PF01172; UPF0023; 1.
CC DR ProDom; PD009796; UPF0023; 1.
CC DR PROSITE; PS01267; UPF0023; 1.
CC KW Hypothetical protein.
CC SEQUENCE 250 AA; 28283 MW; EFC799ADEF73E0E CRC64;
CC
CC Query Match 42.9%; Score 6; DB 1; Length 250;
CC Best Local Similarity 100.0%; Pred. No. 9.6;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC YQ 1 TGLIDP 6
CC DB 215 TGLIDP 220
CC
CC RESULT 5
CC Y237_MYCPN STANDARD; PRT; 294 AA.
CC AC F75455;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE HYPOTHETICAL PROTEIN MG237 HOMOLOG (F10_ORF294).
CC GN MPN330 OR MP506.
CC OS Mycoplasma pneumoniae.
CC OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
CC OC Mycoplasmataceae; Mycoplasma.
CC OX NCBI_TaxID=2104;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=ATCC 29342 / M129;
CC RX MEDLINE=97105885; PubMed=8948633;
CC RA Himmelsreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
CC Herrmann R.;
CC RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
CC pneumoniae."
CC RL Nucleic Acids Res. 24:4420-4449(1996).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC -----
CC YL22_YEAST STANDARD; PRT; 250 AA.
CC AC Q07953;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
CC DE HYPOTHETICAL 28.3 KDA PROTEIN IN PPRI-SNE7 INTERGENIC REGION.
CC GN YL0202C.
CC OS Saccharomyces cerevisiae (Baker's yeast).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CC OX NCBI_TaxID=4932;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA Obermaier B., Piravandi E., Rinke M.;
CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC RN [2]
CC RP SEQUENCE OF 134-250 FROM N.A.
CC RA Vandenbol M., Portetelle D., Hilger F.;
CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC CC -!- SIMILARITY: BELONGS TO THE UPF0023 FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 273194; CAA97545.1; -
CC DR SGD; S0004012; YLR022C.
CC DR InterPro: IPR002140; UPF0023.
CC DR Pfam; PF01172; UPF0023; 1.
CC DR ProDom; PD009796; UPF0023; 1.
CC DR PROSITE; PS01267; UPF0023; 1.
CC KW Hypothetical protein.
CC SEQUENCE 250 AA; 28283 MW; EFC799ADEF73E0E CRC64;
CC
CC Query Match 42.9%; Score 6; DB 1; Length 250;
CC Best Local Similarity 100.0%; Pred. No. 11;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC YQ 6 PGFQGE 11
CC DB 170 PGFQGE 175
CC
CC RESULT 6
CC CORC_HAEIN STANDARD; PRT; 299 AA.
CC ID CORC_HAEIN STANDARD; PRT; 299 AA.
CC AC Q57368;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
CC DE MAGNESIUM AND COBALT EFFLUX PROTEIN CORC.
CC GN CORC OR HI0301.
CC OS Haemophilus influenzae.
CC OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC OC Haemophilus.
CC OX NCBI_TaxID=727;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=RD / KW20 / ATCC 51907;
CC RX MEDLINE=95350630; PubMed=7542800;
CC RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
CC Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
CC McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
CC Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
CC Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
CC Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
CC Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
CC Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
CC Venter J.C.;
CC RT "Whole-genome random sequencing and assembly of Haemophilus
CC influenzae Rd."
CC RL Science 269:496-512(1995).
CC CC -!- FUNCTION: PLAYS A ROLE IN THE TRANSPORT OF MAGNESIUM AND COBALT
CC IONS (BY SIMILARITY).
CC CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY.
CC CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32716; AAC21966.1; -
CC DR TIGR; HI0301; -
CC DR InterPro: IPR000644; CBS.
CC DR Pfam; PF00571; CBS; 2.
CC DR SMART; SM00116; CBS; 2.
CC KW Magnesium; Cobalt; Transport; CBS domain; Repeat; Complete proteome.
CC FT DOMAIN 73 129 CBS 1.
CC FT DOMAIN 138 191 CBS 2.
CC SEQUENCE 299 AA; 34583 MW; 181F455BF4F0213 CRC64;

```

Query Match 42.9% Score 6; DB 1; Length 299;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FOGELK 13
| | | | |
DB 27 FOGELK 32

RESULT 7
CORC_PASMU STANDARD; PRT; 300 AA.
ID CORC_PASMU
AC Q9CM13;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAGNESIUM AND COBALT EFFLUX PROTEIN CORC.
GN CORC OR PM1033;
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21145866; PubMed=11248100;
RX May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.;
FT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: PLAYS A ROLE IN THE TRANSPORT OF MAGNESIUM AND COBALT
IONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UPF0053 FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006144; AAK03117.1; -
CC InterPro; IPR000644; CBS.
CC Pfam; PF00571; CBS; 2.
CC SMART; SM00116; CBS; 2.
DR KW Magnesium; Cobalt; Transport; CBS domain; Repeat; Complete proteome.
FT DOMAIN 75 129
FT DOMAIN 138 191
SQ SEQUENCE 300 AA; 34683 MW; CBB691724B44FF62 CRC64;

Query Match 42.9% Score 6; DB 1; Length 300;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FOGELK 13
| | | | |
DB 29 FOGELK 34

RESULT 8
PEX7_SCHPO STANDARD; PRT; 308 AA.
ID PEX7_SCHPO
AC P78738; O99703;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR (PTS2 RECEPTOR) (PEROXIN-7).
GN PEX7 OR SPAC1834.12.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PR745;
RA Yoshioka S., Kato K., Okayama H.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-24 FROM N.A.
RC STRAIN=972;
RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO THE N-TERMINAL PTS2-TYPE PEROXISOMAL TARGETING
CC SIGNAL AND PLAYS AN ESSENTIAL ROLE IN PEROXISOMAL PROTEIN IMPORT.
CC -1- SUBUNIT: INTERACTS WITH PEX5 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D89147; BAA13809.1; ALT_INIT.
CC EMBL; AL157734; CAB57580.1; -
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 6.
CC PRINTS; PR00320; GPROTEINRPT.
CC SMART; SM00320; WD40; 6.
CC PROSITE; PS00678; WD_REPEATS_1; 1.
CC PROSITE; PS00082; WD_REPEATS_2; 2.
DR KW Peroxisome; Repeat; WD repeat; Transport; Protein transport.
FT REPEAT 57 88
FT REPEAT 101 132
FT REPEAT 145 176
FT REPEAT 187 218
FT REPEAT 231 262
FT REPEAT 274 306
SQ SEQUENCE 308 AA; 34588 MW; B0FAC61A366F5D67 CRC64;

Query Match 42.9% Score 6; DB 1; Length 308;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGFOGE 11
| | | | |
DB 7 PGFOGE 12

RESULT 9
LNT_PSEAE STANDARD; PRT; 511 AA.
ID LNT_PSEAE
AC Q9Z166;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.-) (ALP N-ACYLTRANSFERASE)
DE (COPPER HOMEOSTASIS PROTEIN CUTE HOMOLOG).
GN LNT OR CUTE OR PA3984.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RA Qi H.-Y., Gupta S.D., Wu H.C., Rick P.D.;
RT "Identification of N-acyltransferase (lnt) gene from Pseudomonas

```

RT aeruginosa.";
RN Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RL [2]
RC SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PA01;
RA MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: TRANSFERS THE FATTY ACYL GROUP ON MEMBRANE LIPOPROTEINS
CC (BY SIMILARITY).
CC -!- PATHWAY: LIPOPROTEIN BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE APOLIPOPROTEIN N-ACYLTRANSFERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF038595; AAC97167.1; -.
DR EMBL: AE004816; AAG07371.1; -.
DR InterPro: IPR003010; CN_hydrolase.
DR Pfam: PF00795; CN_hydrolase; 1.
KW Transferase; Acyltransferase; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 125 145 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 482 502 POTENTIAL.
SQ SEQUENCE 511 AA; 56016 MW; 9DAE98422EBFB35 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 511;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGLIDP 6
DB 443 TGLIDP 448

RESULT 10
CA28_HUMAN STANDARD; PRT; 635 AA.
AC P25067;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 2(VIII) CHAIN (ENDOTHELIAL COLLAGEN) (FRAGMENT).
GN COL9A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91210292; PubMed=2019595;
RA Muragaki Y., Jacenko O., Apte S., Mattei M.-G., Ninomiya Y.,

```

```

RA Olsen B.R.;
RT "The alpha 2(VIII) collagen gene. A novel member of the short chain
RT collagen family located on the human chromosome 1.";
RL J. Biol. Chem. 266:7721-7727(1991).
CC -!- FUNCTION: MAJOR COMPONENT OF THE DESCENDING MEMBRANE (BASEMENT
CC MEMBRANE) OF CORNEAL ENDOTHELIAL CELLS.
CC -!- SUBUNIT: MAY FORM HOMOTRIMERS, OR HETEROTRIMERS IN ASSOCIATION
CC WITH ALPHA 1(VIII) TYPE COLLAGENS.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: STRONG, TO ALPHA 1 TYPES VIII AND X COLLAGENS.
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M60832; AAA62822.1; -.
DR MIM: 120252; -.
DR InterPro: IPR001073; Clq.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; Clq; 1.
DR Pfam: PF01391; Collagen; 7.
DR SMART: SM00110; Clq; 1.
DR PROSITE: PS01113; Clq; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen.
FT NON_TER 1 1
FT DOMAIN <1 11 NONHELICAL REGION (NC2).
FT DOMAIN 12 468 TRIPLE-HELICAL REGION.
FT DOMAIN 469 635 NONHELICAL REGION (NC1).
FT DOMAIN 500 635 C1Q.
SQ SEQUENCE 635 AA; 60527 MW; 22A261164754F771 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 635;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGFQGE 11
DB 119 PGFQGE 124

RESULT 11
SIPL_HUMAN STANDARD; PRT; 1214 AA.
AC O60315; O9UED1;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZINC FINGER HOMEBOX PROTEIN 1B (SMAD INTERACTING PROTEIN 1).
GN ZFXH1B OR ZFX1B OR KIAA0569.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [2]
RP SEQUENCE OF 1161-1214 FROM N.A.

```

RC TISSUE=Fetal brain; PubMed=9853615;
RX MEDLINE=99068504; PubMed=9853615;
RA Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.A.;
RT "Selection system for genes encoding nuclear-targeted proteins.";
RL Nat. Biotechnol. 16:1338-1342(1998).
CC -1- FUNCTION: TRANSCRIPTIONAL INHIBITOR THAT BINDS TO DNA SEQUENCE 5'-
CC CACCT-3' IN DIFFERENT PROMOTERS (BY SIMILARITY).
CC -1- SUBUNIT: BINDS ACTIVATED SMAD1, ACTIVATED SMAD2 AND ACTIVATED
CC SMAD3; BINDING WITH SMAD4 IS NOT DETECTED (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMODOMAIN PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AB011141; BAA25495.1; -;
DR ENBL; AB015341; BAA34798.1; -;
DR HSP; P08047; ISP2.
DR InterPro; IPR001336; Homeobox.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR PRINTS; PRO0048; ZINC-FINGER.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf-C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW Homeobox; Repressor; Metal-binding; Repeat.
FT DOMAIN 437 487 SMAD-MH2 BINDING DOMAIN (BY SIMILARITY).
FT ZN_FING 211 234 C2H2-TYPE.
FT ZN_FING 241 263 C2H2-TYPE.
FT ZN_FING 282 304 C2H2-TYPE.
FT ZN_FING 310 334 C2H2-TYPE (ATYPICAL).
FT DNA_BIND 999 1021 HOMEBOX-LIKE.
FT ZN_FING 1027 1049 C2H2-TYPE.
FT ZN_FING 1055 1076 C2H2-TYPE (ATYPICAL).
FT DOMAIN 1084 1214 GLU-RICH (ACIDIC).
SQ SEQUENCE 1214 AA; 136446 MW; B578FD91339C3FDD CRC64;

Query Match 42.9%; Score 6; DB 1; Length 1214;
Best Local Similarity 100.0%; Pred No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFQEL 12
Db 960 GFQEL 965

RESULT 12
ID CA14_HUMAN STANDARD; PRT; 1669 AA.
AC P02462;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN COL4A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89340433; PubMed=2701944;
RA Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;

RT "Structural organization of the gene for the alpha 1 chain of human
RT type IV collagen.";
RL J. Biol. Chem. 264:13565-13571(1989).
RN [2]
RP SEQUENCE OF 46-1257 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88083584; PubMed=3691802;
RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
RT "Complete primary structure of the alpha 1-chain of human basement
RT membrane (type IV) collagen.";
RL FEBS Lett. 225:188-194(1987).
RN [3]
RP SEQUENCE OF 1-943 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88029471; PubMed=3311751;
RA Brazel D., Oberbauer I., Dieringer H., Babel W., Glanville R.W.,
RA Deutzmann R., Kuehn K.;
RT "Completion of the amino acid sequence of the alpha 1 chain of human
RT basement membrane collagen (type IV) reveals 21 non-triplet
RT interruptions located within the collagenous domain.";
RL Eur. J. Biochem. 168:529-536(1987).
RN [4]
RP SEQUENCE OF 28-243.
RX MEDLINE=86004708; PubMed=4043082;
RA Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
RT "Amino acid sequence of the N-terminal aggregation and cross-linking
RT region (7S domain) of the alpha 1 (IV) chain of human basement
RT membrane collagen.";
RL Eur. J. Biochem. 152:213-219(1985).
RN [5]
RP SEQUENCE OF 534-1447.
RX MEDLINE=85003629; PubMed=6434307;
RA Babel W., Glanville R.W.;
RT "Structure of human-basement-membrane (type IV) collagen. Complete
RT amino-acid sequence of a 914-residue-long pepsin fragment from the
RT alpha 1(IV) chain.";
RL Eur. J. Biochem. 143:545-556(1984).
RN [6]
RP SEQUENCE OF 1256-1669 FROM N.A.
RX MEDLINE=85027819; PubMed=2581969;
RA Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
RA Cheung M.-C., Prockop D.J., Boyd C.D.;
RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
RT procollagen reveal an unusual homology of amino acid sequences in two
RT halves of the carboxyl-terminal domain.";
RL J. Biol. Chem. 260:7681-7687(1985).
RN [7]
RP SEQUENCE OF 1259-1669 FROM N.A.
RX MEDLINE=85216555; PubMed=2582422;
RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
RA Refalides N.A., Myers J.C.;
RT "Restricted homology between human alpha 1 type IV and other
RT procollagen chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [9]
RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'

MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
NIDODEN.
-1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
-1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
TRIPLE-HELICAL 7S DOMAIN.
-1- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.
-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
-1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
IV COLLAGENS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M26576; AAA53098.1; JOINED.
EMBL; J04217; AAA53098.1; JOINED.
EMBL; M26550; AAA53098.1; JOINED.
EMBL; M26540; AAA53098.1; JOINED.
EMBL; M26542; AAA53098.1; JOINED.
EMBL; M26543; AAA53098.1; JOINED.
EMBL; M26544; AAA53098.1; JOINED.
EMBL; M26545; AAA53098.1; JOINED.
EMBL; M26546; AAA53098.1; JOINED.
EMBL; M26547; AAA53098.1; JOINED.
EMBL; M26537; AAA53098.1; JOINED.
EMBL; M26538; AAA53098.1; JOINED.
EMBL; M26548; AAA53098.1; JOINED.
EMBL; M26549; AAA53098.1; JOINED.
EMBL; M26551; AAA53098.1; JOINED.
EMBL; M26552; AAA53098.1; JOINED.
EMBL; M26553; AAA53098.1; JOINED.
EMBL; M26554; AAA53098.1; JOINED.
EMBL; M26555; AAA53098.1; JOINED.
EMBL; M26556; AAA53098.1; JOINED.
EMBL; M26557; AAA53098.1; JOINED.
EMBL; M26539; AAA53098.1; JOINED.
EMBL; M26558; AAA53098.1; JOINED.
EMBL; M26559; AAA53098.1; JOINED.
EMBL; M26560; AAA53098.1; JOINED.
EMBL; M26561; AAA53098.1; JOINED.
EMBL; M26562; AAA53098.1; JOINED.
EMBL; M26536; AAA53098.1; JOINED.
EMBL; M26563; AAA53098.1; JOINED.
EMBL; M26541; AAA53098.1; JOINED.
EMBL; M26564; AAA53098.1; JOINED.
EMBL; M26565; AAA53098.1; JOINED.
EMBL; M26566; AAA53098.1; JOINED.
EMBL; M26567; AAA53098.1; JOINED.
EMBL; M26568; AAA53098.1; JOINED.
EMBL; M26569; AAA53098.1; JOINED.
EMBL; M26570; AAA53098.1; JOINED.
EMBL; M26571; AAA53098.1; JOINED.
EMBL; M26572; AAA53098.1; JOINED.
EMBL; M26573; AAA53098.1; JOINED.
EMBL; M26574; AAA53098.1; JOINED.
EMBL; M26575; AAA53098.1; JOINED.
EMBL; Y00706; CAA68698.1; JOINED.
EMBL; X05561; CAA29075.1; JOINED.
EMBL; M10940; AAA52006.1; JOINED.
EMBL; M11315; AAA52042.1; JOINED.

DR PIR; S16876; CGHU4B.
DR MIM; 120130; .
DR InterPro; IPR001442; C4.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
DR ProDom; PD003923; C4; 2.
DR SMART; SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Basement membrane;
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN 174 1440 TRIPLE-HELICAL REGION.
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1662.
FT DISULFID 1570 1665 OR 1665.
FT DISULFID 1604 1662 OR 1665.
FT DISULFID 1616 1622 OR 1665.
FT CONFLICT 237 238 SG -> KE (IN REF. 4).
FT CONFLICT 241 241 G -> K (IN REF. 4).
FT CONFLICT 319 319 Q -> A (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 2).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
SQ SEQUENCE 1669 AA; 160611 MW; 3BBA6DFFB9B8A4 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 1669;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGFQG 10
| | | | |
Db 1285 DPGFQG 1290

RESULT 13
CA14_MOUSE
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogel G.;
RT "cDNA clones completing the nucleotide and derived amino acid
sequence of the alpha 1 chain of basement membrane (type IV) collagen

RT from mouse.";

RL [3] SEQUENCE OF 1149-1424 FROM N.A.

RP MEDLINE=86301886; PubMed=3755692;

RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;

RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a

RL synthetic oligodeoxynucleotide.";

RL Gene 43:301-304(1986).

RR [4]

RR SEQUENCE OF 1276-1669 FROM N.A.

RR MEDLINE=85127033; PubMed=2578961;

RR Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,

RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;

RT "Amino acid sequence of the non-collagenous globular domain (NC1) of

RL the alpha 1(IV) chain of basement membrane collagen as derived from

RL complementary DNA.";

RL Eur. J. Biochem. 147:217-224(1985).

RR [5]

RR SEQUENCE OF 1441-1669 FROM N.A.

RR MEDLINE=87250460; PubMed=3597383;

RR Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,

RA Saus J., Pihlajaniemi T.;

RT "Extensive homology between the carboxyl-terminal peptides of mouse

RL alpha 1(IV) and alpha 2(IV) collagen.";

RL J. Biol. Chem. 262:8496-8499(1987).

RR [6]

RR PARTIAL SEQUENCE FROM N.A.

RR MEDLINE=86196099; PubMed=3009468;

RR Sakurai Y., Sullivan M., Yamada Y.;

RT "Alpha 1 type IV collagen gene evolved differently from fibrillar

RL collagen genes.";

RL J. Biol. Chem. 261:6654-6657(1986).

RR [7]

RR SEQUENCE OF 1-28 FROM N.A.

RR MEDLINE=89067738; PubMed=3198626;

RR Kaytes P., Wood L., Theriault N.;

RT "Head-to-head arrangement of murine type IV collagen genes.";

RL J. Biol. Chem. 263:19274-19277(1988).

RR [8]

RR SEQUENCE OF 1-28 FROM N.A.

RR MEDLINE=89071759; PubMed=3200851;

RR Burbelo P.D., Martin G.R., Yamada Y.;

RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a

RL bidirectional promoter and a shared enhancer.";

RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).

RR [9]

RR SEQUENCE OF 1-129 FROM N.A.

RR MEDLINE=88243724; PubMed=3379041;

RR Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;

RA "Structure of the amino-terminal portion of the murine alpha 1(IV)

RL collagen chain and the corresponding region of the gene.";

RL J. Biol. Chem. 263:8706-8709(1988).

CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF

CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'

CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
NIDOGEN.

CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-

CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE

CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS

CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE

CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY

CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL

CC TRIPLE-HELICAL 7S DOMAIN.

CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING

CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH

CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF

CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE

CC IV COLLAGENS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL; J03758; AAA37439.1; -

DR EMBL; M23333; AAA51625.1; -

DR EMBL; J04694; AAA50292.1; -

DR EMBL; X06777; CAA29946.1; -

DR EMBL; X02201; CAA26132.1; -

DR EMBL; M15832; AAA37340.1; -

DR EMBL; M14042; AAA37342.1; -

DR EMBL; M12879; AAA37343.1; -

DR EMBL; M13024; -; NOT_ANNOTATED_CDS.

DR EMBL; M13025; -; NOT_ANNOTATED_CDS.

DR EMBL; M13026; AAA37344.1; -

DR EMBL; M13027; AAA37345.1; -

DR EMBL; M13043; AAA37346.1; -

DR EMBL; J04448; AAA37437.1; -

DR PIR; A33525; CGMS4B.

DR MGD; MGI:88454; Col4a1.

DR InterPro: IPR001442; C4.

DR InterPro: IPR000087; Collagen.

DR Pfam: PF01413; C4; 2.

DR Pfam: PF01391; Collagen; 21.

DR ProDom: PD003923; C4; 2.

DR SMART; SM00111; C4; 2.

KW Extracellular matrix; Connective tissue; Basement membrane;

KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.

FT SIGNAL 27

FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.

FT DOMAIN 173 1669 TRIPLE-HELICAL REGION.

FT DISULFID 1441 1669 NONHELICAL REGION (NC1).

FT DISULFID 1460 1551 OR 1546 (BY SIMILARITY).

FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).

FT DISULFID 1505 1511 BY SIMILARITY.

FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).

FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).

FT DISULFID 1616 1622 BY SIMILARITY.

FT CARBOHYD 126 126 N-LINKED (GLCNAC...) (POTENTIAL).

FT CONFLICT 26 26 A -> P (IN REF. 2).

FT CONFLICT 186 186 S -> L (IN REF. 2).

FT CONFLICT 319 319 Q -> S (IN REF. 2).

FT CONFLICT 369 369 Q -> L (IN REF. 2).

FT CONFLICT 403 403 L -> F (IN REF. 2).

FT CONFLICT 481 481 P -> L (IN REF. 2).

FT CONFLICT 493 493 Q -> H (IN REF. 2).

FT CONFLICT 712 712 S -> I (IN REF. 2).

FT CONFLICT 813 813 E -> Q (IN REF. 2).

FT CONFLICT 982 982 Q -> H (IN REF. 2).

FT CONFLICT 1397 1397 V -> S (IN REF. 3).

SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 1669;

Best Local Similarity 100.0%; Pred. No. 54;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGFQG 10

DB 1285 DPGFQG 1290

|||||

RESULT 14

ID CA34_HUMAN STANDARD; PRT; 1670 AA.

AC Q01955;

DC 01-OCT-1996 (Rel. 34, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR (GOODPASTURE ANTIGEN).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94364994; PubMed=8083201;
RA Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.
RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in
RT human tissues.";
RL J. Biol. Chem. 269:23013-23017(1994).
RN [2]
RP REVISIONS.
RA Leinonen A.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. Biol. Chem. 267:19780-19784(1992).
RN [4]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1892840;
RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
RN [5]
RP SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA Pusey C.D.;
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
RT to be the alpha 3 chain of type IV collagen.";
RL J. Clin. Invest. 89:592-601(1992).
RN [6]
RP SEQUENCE OF 1644-1670 FROM N.A.
RC TISSUE=Kidney;
RX Ding J.;
RA Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE=Kidney;
RX MEDLINE=94124597; PubMed=8294492;
RA Feng L., Xia Y., Wilson C.B.;
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions.";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [8]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [9]
RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced.";
RL J. Biol. Chem. 268:12090-12094(1993).

RN [10]
RP VARIANT PRO-1474.
RX MEDLINE=95078827; PubMed=7987301;
RA Lemlink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Reiders S.T., Smeets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
RT recessive Alport syndrome.";
RL Hum. Mol. Genet. 3:1269-1273(1994).
RN [11]
RP VARIANTS AS, AND VARIANTS.
RX MEDLINE=21064696; PubMed=11134255;
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome.";
RL J. Am. Soc. Nephrol. 12:97-106(2001).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
CC NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/V AND
CC 3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
CC C-TERMINAL NC1 DOMAINS.
CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC COCHLEA, LUNG AND BRAIN.
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL
CC N-LINKED GLYCOSYLATION SITE.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -!- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
CC -!- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
CC MALES AND FEMALES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80031; CAA56335.1; -
DR EMBL; M92993; AAA21610.1; -
DR EMBL; S55790; AAB19637.1; -
DR EMBL; M81379; AAA51536.1; -
DR EMBL; L08650; AAA52044.1; -
DR EMBL; U02519; AAA18942.1; -
DR EMBL; U02520; AAA18943.1; -
DR EMBL; AB008495; BAA25064.1; -
DR MIM; 120070; -
DR MIM; 203780; -
DR MIM; 233450; -
DR InterPro; IPR001442; C4.

CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23334; AAA51626.1; -
DR EMBL; M23333; AAA51626.1; JOINED.
DR EMBL; J04695; AAA50293.1; -
DR EMBL; J04448; AAA37438.1; -
DR EMBL; J04647; CAA28308.1; -
DR EMBL; M15833; AAA37341.1; -
DR EMBL; X04410; CAA27998.1; -
DR EMBL; X02896; CAA26655.1; -
DR EMBL; X02897; CAB51614.1; -
DR EMBL; X02898; CAA26657.1; -
DR EMBL; X02899; CAA26658.1; -
DR EMBL; M23334; AAA51627.1; -
DR PIR; A33526; A33526.
DR HSSP; P19972; IKVD.
DR MGD; MGI:88455; Col4a2.
DR InterPro; IPR001442; C4.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 19.
DR ProDom; PD003923; C4; 2.
DR SMART; SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Signal.
FT SIGNAL 1 25
FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
FT CHAIN 184 1707 COLLAGEN ALPHA 2(IV) CHAIN.
FT DOMAIN 184 1479 TRIPLE-HELICAL REGION.
FT DOMAIN 1480 1707 NONHELICAL REGION (NC1).
FT DISULFID 1499 1588 OR 1585 (BY SIMILARITY).
FT DISULFID 1532 1585 OR 1588 (BY SIMILARITY).
FT DISULFID 1544 1550 BY SIMILARITY.
FT DISULFID 1607 1703 OR 1700 (BY SIMILARITY).
FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).
FT DISULFID 1653 1660 BY SIMILARITY.
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1270 1270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1051 1051 P -> R (IN REF. 6).
FT CONFLICT 1097 1097 S -> G (IN REF. 7).
FT CONFLICT 1171 1171 G -> S (IN REF. 6).
FT CONFLICT 1179 1179 P -> R (IN REF. 6).
FT CONFLICT 1241 1241 Q -> E (IN REF. 6).
FT CONFLICT 1328 1328 P -> A (IN REF. 6).
FT CONFLICT 1573 1573 V -> L (IN REF. 4).
FT CONFLICT 1623 1623 Y -> H (IN REF. 4).
SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605FD508 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 1707;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 DPGFQG 10
Db 362 DPGFQG 367
|||||

Search completed: January 31, 2002, 13:39:26
Job time: 86 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:24 : Search time 46.78 seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-80

Perfect score: 14

Sequence: 1 AGVIDEDYRGNGV 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	159	1 DUT_ORFN2	P14597 orf virus (
2	14	100.0	205	1 DUT_RAT	P70583 rattus norv
3	14	100.0	252	1 DUT_HUMAN	P33316 homo sapien
4	10	71.4	144	1 DUT_VACCV	P17374 vaccinia vi
5	10	71.4	147	1 DUT_VACCC	P21035 vaccinia vi
6	10	71.4	147	1 DUT_VARV	P33826 variola vir
7	7	50.0	1003	1 MYSE_DICDI	Q03479 dictyostell
8	7	50.0	2021	1 OMPA_RICCN	Q52657 rickettsia
9	6	42.9	171	1 PSAN_ARATH	P49107 arabidopsis
10	6	42.9	178	1 DUT_ADEG1	Q89662 avian adeno
11	6	42.9	218	1 IM23_SCHMA	P19331 schistosoma
12	6	42.9	271	1 DAPB_MYCBO	P46829 mycobacteri
13	6	42.9	289	1 YHDE_BACSU	O07575 bacillus su
14	6	42.9	624	1 SERA_ARATH	O04130 arabidopsis
15	6	42.9	757	1 RRPI_IAPOM	Q82571 influenza a
16	6	42.9	766	1 EYA_DROME	Q05201 drosophila
17	6	42.9	878	1 SYA_BUCAI	P57483 buchnera ap
18	6	42.9	1365	1 YAKI_SCHPO	O09915 schizosacch
19	6	42.9	1391	1 N157_YEAST	P40064 saccharomyc
20	5	35.7	76	1 VAI9_VARV	P33842 variola vir
21	5	35.7	82	1 COAB_BPPF1	P03621 bacterioph
22	5	35.7	93	1 RLAQ_PIG	Q29214 sus scrofa
23	5	35.7	100	1 CHA3_BOMMO	P08929 bombyx mori
24	5	35.7	103	1 VP16_NPVOP	P24079 orgyia pseu
25	5	35.7	109	1 YS88_CAEEL	Q09384 caenorhabdi
26	5	35.7	114	1 UNR_CAVPO	P29174 cavia porce
27	5	35.7	120	1 YHFY_ECOLI	P45551 escherichia
28	5	35.7	129	1 CHA1_BOMMO	P08826 bombyx mori
29	5	35.7	131	1 YD2E_SCHPO	Q10262 schizosacch
30	5	35.7	132	1 CHA2_BOMMO	P08825 bombyx mori
31	5	35.7	140	1 DUT_SCHPO	Q9p6q5 schizosacch
32	5	35.7	141	1 DUT_CHVP1	O41033 paramecium
33	5	35.7	143	1 RK2_SOYBN	P18663 glycine max

O48500 bacterioph
P43580 saccharomyc
P38839 saccharomyc
P16792 human cytom
Q03740 mus musculu
P43058 candida alb
Q9YYS0 avian adeno
Q9ZK53 helicobacte
O25776 helicobacte
P32518 lycopersico
P08209 bos taurus
P07320 homo sapien

148 1 DUT_BPT5
148 1 YFB2_YEAST
148 1 YHT6_YEAST
157 1 UL51_HCMVA
158 1 CRGF_MOUSE
159 1 DUT_CANAL
163 1 DUT_ADEG8
164 1 FLAV_HELPJ
164 1 FLAV_HELPY
169 1 DUT_LYCES
173 1 CRGD_BOVIN
173 1 CRGD_HUMAN

ALIGNMENTS

```

RESULT 1
DUT_ORFN2
ID DUT_ORFN2 STANDARD; PRT; 159 AA.
AC P14597;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPASE) (DUTP PYROPHOSPHATASE).
OS Orf virus (strain NZ2) (OV NZ-2).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Parapoxvirus.
OX NCBI_Taxid=10259;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90021200; PubMed=2678731;
RA Mercer A.A., Fraser K.M., Stockwell P.A., Robinson A.J.;
RT "A homologue of retroviral pseudoproteases in the parapoxvirus, orf
virus.";
RL Virology 172:665-668(1989).
CC -!- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
CC URACIL CANNOT BE INCORPORATED INTO DNA.
CC -!- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PROTEASE-LIKE PROTEIN
CC (PSEUDOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30023; AAA46786.1; -
CC DR HSSP; P06968; 1DUU.
CC DR InterPro: IPR001428; dUTPase.
CC DR Pfam: PF00692; dUTPase; 1.
CC KW Hydrolase; Nucleotide metabolism.
CC SEQUENCE 159 AA; 16893 MW; 08CD852D47AE17AD CRC64;

```

Query Match 100.0%; Score 14; DB 1; Length 159;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNGV 14

Db 81 AGVIDEDYRGNGV 94

RESULT 2

DUT_RAT

ID DUT_RAT STANDARD; PRT; 205 AA.
 AC P70583;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
 DE (DUTPASE) (DUTP PYROPHOSPHATASE) (PPAR-INTERACTING PROTEIN 4). (PIP4).
 GN DUT.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=97066956; PubMed=8910358;
 RA Chu R.Y., Lin Y.L., Rao M.S., Reddy J.K.;
 RT "Cloning and identification of rat deoxyuridine triphosphatase as an
 RT inhibitor of peroxisome proliferator-activated receptor alpha.";
 RL J. Biol. Chem. 271:27670-27676(1996).
 RN [2]
 RP REVISIONS.
 RA Chu R.Y., Lin Y.L., Rao M.S., Reddy J.K.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 CC URACIL CANNOT BE INCORPORATED INTO DNA. INHIBITS PEROXISOME
 CC PROLIFERATOR-ACTIVATED RECEPTOR (PPAR) ACTIVITY BY BINDING OF ITS
 CC RETINOID X RECEPTOR.
 CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC -1- COFACTOR: MAGNESIUM (BY SIMILARITY).
 CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. BINDING TO PPAR INDUCES
 CC TRANSLOCATION TO THE NUCLEUS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES EXAMINED. HIGHER
 CC LEVELS IN HEART AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U64030; AAC34734.2; -
 CC HSSP: P16088; 1DUT.
 DR InterPro: IPR001428; dutpase.
 DR Pfam: PF00692; dutpase; 1.
 KW Hydrolase; Nucleotide metabolism; Magnesium.
 SQ SEQUENCE 205 AA; 27003 MW; A9D54EBF5ED015C4 CRC64;
 Query Match 100.0%; Score 14; DB 1; Length 205;
 Best Local Similarity 100.0%; Pred. No. 3.6e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGVIDEYRGNGV 14
 Db 139 AGVIDEYRGNGV 152
 RESULT 3
 ID DUT_HUMAN STANDARD; PRT; 252 AA.
 AC P33316; Q16860; Q16708; O14785;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE PRECURSOR

DE GN (EC 3.6.1.23) (DUTPASE) (DUTP PYROPHOSPHATASE).
 DUT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (DUT-M AND DUT-N).
 RA Pearlman R.E.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (DUT-M).
 RA Ladner R.D., Caradonna S.J.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (DUT-N).
 RA Cohen D., Heng H.H.Q., Shi X.M., McIntosh E.M., Tsui L.C.,
 RA Pearlman R.E.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (DUT-N).
 RA TISSUE=T-cell;
 RX MEDLINE=96205967; PubMed=8631816;
 RA Ladner R.D., McNulty D.E., Carr S.A., Roberts G.D., Caradonna S.J.;
 RT "Characterization of distinct nuclear and mitochondrial forms of
 RT human deoxyuridine triphosphate nucleotidohydrolase.";
 RL J. Biol. Chem. 271:7745-7751(1996).
 RN [5]
 RP SEQUENCE OF 112-252 FROM N.A.
 RX MEDLINE=92390380; PubMed=1325640;
 RA McIntosh E.M., Ager D.D., Gadsden M.H., Haynes R.H.;
 RT "Human dUTP pyrophosphatase: cDNA sequence and potential biological
 RT importance of the enzyme.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8020-8024(1992).
 RN [6]
 RP ERRATUM.
 RA McIntosh E.M., Ager D.D., Gadsden M.H., Haynes R.H.;
 RL Proc. Natl. Acad. Sci. U.S.A. 90:4328-4328(1993).
 RN [7]
 RP SEQUENCE OF 112-252 FROM N.A., PARTIAL SEQUENCE, AND PHOSPHORYLATION.
 RC TISSUE=Lymphocytes;
 RX MEDLINE=93281681; PubMed=8389461;
 RA Strahler J.R., Zhu X.-X., Wang Y.K., Hora N., Andrews P.C.,
 RA Roseman N.A., Neel J.V., Turka L., Hanash S.M.;
 RT "Maturation stage and proliferation-dependent expression of dUTPase
 RT in human T cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:4991-4995(1993).
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 CC URACIL CANNOT BE INCORPORATED INTO DNA.
 CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC -1- COFACTOR: MAGNESIUM.
 CC -1- ENZYME REGULATION: PHOSPHORYLATION IS NECESSARY FOR ACTIVITY.
 CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
 CC -1- SUBCELLULAR LOCATION: DUT-N IS NUCLEAR, DUT-M IS MITOCHONDRIAL.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DUT-M (SHOWN HERE) AND DUT-N;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. DUT-N IS THE MORE ABUNDANTLY
 CC EXPRESSED FORM.
 CC -1- TISSUE SPECIFICITY: FOUND IN A VARIETY OF TISSUES.
 CC -1- PTM: PHOSPHORYLATION IN MATURE T-CELLS OCCUR IN A CELL CYCLE-
 CC DEPENDENT MANNER.
 CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF018432; AAB71393.1; -


```
CC CC      URACIL CANNOT BE INCORPORATED INTO DNA.
CC CC      -1- CATALYTIC ACTIVITY: DUTP + H(2)O -> DUMP + PYROPHOSPHATE.
CC CC      -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC CC      -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PROTEASE-LIKE PROTEIN
CC CC      (PSEUDOPROTEASE).
CC CC      -----
CC CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL; M35027; AAA48015.1; -
CC CC      EMBL; M57977; AAA48296.1; -
CC CC      PIR; G42506; PRV27F.
CC CC      HSP; P16088; IDUT.
CC CC      InterPro: IPR001428; dUTPase.
CC CC      Pfam: PF00692; dUTPase; 1.
CC CC      Hydrolase; Nucleotide metabolism.
CC CC      SQ SEQUENCE 147 AA; 16264 MW; DA961240BF85FA2A CRC64;
CC CC      -----
CC CC      Query Match 71.4%; Score 10; DB 1; Length 147;
CC CC      Best Local Similarity 100.0%; Pred. No. 0.00042;
CC CC      Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC      -----
CC CC      QY 2 GVIDEDYRGN 11
CC CC      Db 82 GVIDEDYRGN 91
CC CC      -----
CC CC      RESULT 6
CC CC      ID DUT_VARY STANDARD; PRT; 147 AA.
CC CC      AC P33826; 1994 (Rel. 28, Created)
CC CC      DT 01-FEB-1994 (Rel. 28, Last sequence update)
CC CC      DT 15-FEB-1998 (Rel. 37, Last annotation update)
CC CC      DE DEXYRIDINE 5'-TRIPHOSPHATE NUCLEOTIDHYDROLASE (EC 3.6.1.23)
CC CC      DE (DUTPASE) (DUTP PYROPHOSPHATASE).
CC CC      GN F2L OR C6L OR E2L.
CC CC      OS Variola virus.
CC CC      OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
CC CC      OC Orthopoxvirus.
CC CC      OX NCBI_TaxID=10255;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RC STRAIN-INDIA-1967 / ISOLATE IND3;
CC CC      RX MEDLINE=93202281; PubMed=8384129;
CC CC      RA Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
CC CC      RT "Genes of variola and vaccinia viruses necessary to overcome the host
CC CC      RT protective mechanisms.";
CC CC      RL FEBS Lett. 319:80-83(1993).
CC CC      [2]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RC STRAIN-BANGLADESH-1975;
CC CC      RX MEDLINE=94088747; PubMed=8264798;
CC CC      RA Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
CC CC      RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
CC CC      RA Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
CC CC      RA Venter C.J.;
CC CC      RT "Potential virulence determinants in terminal regions of variola
CC CC      RT smallpox virus genome.";
CC CC      RL Nature 366:748-751(1993).
CC CC      [3]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RC STRAIN-SOMALIA-1977, CONGO-1965, AND GARCIA-1966;
CC CC      RA Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
CC CC      RA Taimen A.V., Shchelkunov S.N., Esposito J.J.;
CC CC      RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC CC      -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
```

```
CC CC      PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC CC      AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
CC CC      URACIL CANNOT BE INCORPORATED INTO DNA.
CC CC      -1- CATALYTIC ACTIVITY: DUTP + H(2)O -> DUMP + PYROPHOSPHATE.
CC CC      -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC CC      -----
CC CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL; X69198; CAA48967.1; -
CC CC      EMBL; L22579; AAA60774.1; -
CC CC      EMBL; U18340; AAA69437.1; -
CC CC      EMBL; U18337; AAA69331.1; -
CC CC      EMBL; U18338; AAA69372.1; -
CC CC      PIR; F36839; F36839.
CC CC      HSP; P16088; IDUT.
CC CC      InterPro: IPR001428; dUTPase.
CC CC      Pfam: PF00692; dUTPase; 1.
CC CC      Hydrolase; Nucleotide metabolism.
CC CC      SQ SEQUENCE 147 AA; 16464 MW; D918ED4F482ECDAL CRC64;
CC CC      -----
CC CC      Query Match 71.4%; Score 10; DB 1; Length 147;
CC CC      Best Local Similarity 100.0%; Pred. No. 0.00042;
CC CC      Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC      -----
CC CC      QY 2 GVIDEDYRGN 11
CC CC      Db 82 GVIDEDYRGN 91
CC CC      -----
CC CC      RESULT 7
CC CC      ID MISEIDCDI STANDARD; PRT; 1003 AA.
CC CC      AC Q03479; 1994 (Rel. 28, Created)
CC CC      DT 01-FEB-1994 (Rel. 28, Last sequence update)
CC CC      DT 01-FEB-1994 (Rel. 28, Last sequence update)
CC CC      DT 30-MAY-2000 (Rel. 38, Last annotation update)
CC CC      DE MYOSIN IE HEAVY CHAIN.
CC CC      GN MYOE OR DMIE.
CC CC      OS Dictyostelium discoideum (Slime mold).
CC CC      OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
CC CC      OX NCBI_TaxID=44689;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RC STRAIN-AX3;
CC CC      RX MEDLINE=93277957; PubMed=8504170;
CC CC      RA Urrutia R., Jung G., Hammer J.A. III;
CC CC      RT "The Dictyostelium myosin IE heavy chain gene encodes a truncated
CC CC      RT isoform that lacks sequences corresponding to the actin binding site
CC CC      RT in the tail.";
CC CC      RL Blochim. Biophys. Acta 1173:225-229(1993).
CC CC      -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC CC      ACTIVITY THAT IS ACTIVATED BY ACTIN. MAY PLAY A ROLE IN MOVING
CC CC      MEMBRANES RELATIVE TO ACTIN
CC CC      -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC CC      AND A LIGHT CHAIN INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC CC      -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC CC      -1- SIMILARITY: CONTAINS 2 IO DOMAINS.
CC CC      -----
CC CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
```

```

DR EMBL; L06805; AAA33201.1; -.
DR PIR; S33760; S33760.
DR HSSP; P08799; 1MND.
DR DictyDb; DD01049; myoe.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR ProDom; PD000355; myosin_HEAVY.
DR PRINTS; PR00193; MYOSINHEAVY.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 2.
DR Myosin; Actin-binding; ATP-binding; Calmodulin-binding; Repeat;
KW Multigene family.
FT DOMAIN 1 693 MYOSIN HEAD-LIKE.
FT DOMAIN 694 722 IQ 1.
FT DOMAIN 716 745 IQ 2.
FT DOMAIN 796 1003 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
FT NP_BIND 101 108 ATP (BY SIMILARITY).
FT DOMAIN 556 630 ACTIN-BINDING.
FT DOMAIN 858 958 MEMBRANE-BINDING (POTENTIAL).
SO SEQUENCE 1003 AA; 114852 MW; B6E758BEC035766F CRC64;

Query Match 50.0%; Score 7; DB 1; Length 1003;
Best Local Similarity 100.0%; Pred. No. 3 3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGVIDED 7
Db 594 AGVIDED 600

RESULT 8
ID OMPA_RICCN STANDARD; PRT; 2021 AA.
AC Q52657; P95591; P95592; P95593; P95594; Q52667; Q52668; Q52669; .
AC Q52670; Q52674;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN A PRECURSOR (190 KDA ANTIGEN) (CELL SURFACE
DE ANTIGEN) (ROMPA) (ROMP A).
GN OMPA OR RC1273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kda antigen-encoding gene of Rickettsia
RT conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Rickettsia conorii strain Malish 7 complete genome.";
RL Submitted (JUN-2001) to the SWISS-PROT data bank.
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.-E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein rompA.";

```

```

RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.-E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein rompA.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- PTM: GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U01028; AAA17405.1; -.
DR EMBL; U43794; AAB49549.1; -.
DR EMBL; U43798; AAB49550.1; -.
DR EMBL; U43806; AAB49551.1; -.
DR EMBL; U45244; AAB49566.1; -.
DR EMBL; U45918; AAA86663.1; -.
DR EMBL; U83440; AAC35176.1; -.
DR EMBL; U83443; AAC35179.1; -.
DR EMBL; U83448; AAC35184.1; -.
DR EMBL; U83453; AAC35189.1; -.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein.
FT CHAIN 1 38 POTENTIAL.
FT DOMAIN 39 2021 OUTER MEMBRANE PROTEIN A.
FT DOMAIN 238 946 THR-RICH.
FT DOMAIN 1424 1528
FT VARIANT 60 60 N -> NN (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 76 76 R -> H (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 86 137 MISSING (IN STRAIN M1).
FT VARIANT 126 133 MISSING (IN STRAIN MOROCCAN).
FT VARIANT 953 954 VT -> II (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 1245 1245 D -> A (IN STRAINS INDIAN TICK TYPHUS, M1
AND MOROCCAN).
FT VARIANT 1308 1308 N -> H (IN STRAIN MOROCCAN).
FT VARIANT 1877 1877 M -> I (IN STRAIN INDIAN TICK TYPHUS).
FT CONFLICT 10 10 Q -> K (IN REF. 1).
FT CONFLICT 92 92 I -> V (IN REF. 1).
FT CONFLICT 126 126 V -> I (IN REF. 1).
FT CONFLICT 137 137 T -> N (IN REF. 1).
FT CONFLICT 157 157 G -> D (IN REF. 1).
FT CONFLICT 368 369 IS -> VN (IN REF. 1).
FT CONFLICT 374 388 KATLGAIKATTTK -> LLQVGQGVKANTIN (IN
REF. 1).
FT CONFLICT 640 640 N -> D (IN REF. 1).
FT CONFLICT 669 669 V -> I (IN REF. 1).
FT CONFLICT 793 793 N -> D (IN REF. 1).
FT CONFLICT 804 804 VN -> IS (IN REF. 1).
FT CONFLICT 809 823 LLRVQGVKSNNTIN -> KATLGAIKATTTK (IN
REF. 1).
FT CONFLICT 898 898 D -> Y (IN REF. 1).
FT CONFLICT 908 908 P -> N (IN REF. 1).
FT CONFLICT 985 985 N -> K (IN REF. 1).
FT CONFLICT 1009 1009 L -> S (IN REF. 1).
FT CONFLICT 1013 1013 Y -> S (IN REF. 1).
FT CONFLICT 1182 1182 K -> Q (IN REF. 1).
FT CONFLICT 1314 1314 N -> Y (IN REF. 4).
FT CONFLICT 1451 1451 H -> N (IN REF. 1).
FT CONFLICT 1624 1624 G -> D (IN REF. 1).
FT CONFLICT 1628 1628 E -> G (IN REF. 1).
FT CONFLICT 1872 1872 A -> V (IN REF. 1).

```

[illegible]

<hr/>					
Query Match	50.0%; Score 7; DB 1; Length 2021;				
Best Local Similarity	100.0%; Pred. No. 6.2;				
Matches	7; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
<hr/>					
OY	1 AGVIDED 7				
Db	1623 AGVIDED 1629				
<hr/>					
RESULT	9				
PSAN_ARATH					
ID	PSAN_ARATH	STANDARD;	PRT;	171 AA.	
AC	P49107;				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	PHOTOSYSTEM I REACTION CENTRE SUBUNIT N PRECURSOR (PSI-N).				
GN	PSAN OR ATSG64040 OR MHJ24.2.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;				
OX	eurosid II; Brassicales; Brassicaceae; Arabidopsi.				
NCBI_TaxID=3702;					
[1]					
SEQUENCE FROM N.A.					
RA	Sehnke P.C., Ferl R.J.;				
RT	"Nucleotide sequence of an Arabidopsis thaliana cDNA clone encoding				
RT	the complete precursor for a homolog to the barley extrinsic thylakoid				
RT	luminal polypeptide PST-N".				
(In) Plant Gene Register	PCR95-088.				
[2]					
SEQUENCE FROM N.A.					
RC	STRAIN-CV COLUMBIA;				
RX	MEDLINE=98162728; Pubmed=9501997;				
RA	Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,				
RA	Tabata S.;				
RT	Structural analysis of Arabidopsis thaliana chromosome 5. III.				
RT	Sequence features of the regions of 1,191,918 bp covered by seventeen				
RT	physically assigned pl clones. ;				
DNA Res. 4:401-414(1997).					
-!- FUNCTION: MAY FUNCTION IN MEDIATING THE BINDING OF THE ANTENNA					
CC	COMPLEXES TO THE PSI REACTION CENTRE AND CORE ANTENNA.				
-!- SUBCELLULAR LOCATION: ON THE LUMINAL SURFACE OF THE THYLAKOID					
MEMBRANE.					
-!- SIMILARITY: BELONGS TO THE PSAN FAMILY.					

This SWISS-PROT entry is copyright. It is produced through a collaboration					
between the Swiss Institute of Bioinformatics and the EMBL outstation -					
the European Bioinformatics Institute. There are no restrictions on its					
use by non-profit institutions as long as its content is in no way					
modified and this statement is not removed. Usage by and for commercial					
entities requires a license agreement (See http://www.isb-sib.ch/announcement/					
FT	SEQUENCE 178 AA; 19219 MW; 477C885A94AD16F1 CRC64;				
S -> F (IN REF. 1; AAB26434).					
477C885A94AD16F1 CRC64;					

EMBL: Z17216; CAAT78921.1; -					
DR	EMBL: S61107; AAB26434.1; -				
DR	EMBL: U46933; AAC54895.1; -				
DR	HSSP: P06968; IDUD.				
DR	InterPro: IPRO01428; dutpase.				
DR	Pfam: PF00692; dutpase; 1.				
KW	Hydrolase; Nucleotide metabolism.				
CONFLICT	156 156				
FT	SEQUENCE 178 AA; 19219 MW; 477C885A94AD16F1 CRC64;				
S -> F (IN REF. 1; AAB26434).					
477C885A94AD16F1 CRC64;					

EMBL: Z17216; CAAT78921.1; -					
DR	EMBL: S61107; AAB26434.1; -				
DR	EMBL: U46933; AAC54895.1; -				
DR	HSSP: P06968; IDUD.				
DR	InterPro: IPRO01428; dutpase.				
DR	Pfam: PF00692; dutpase; 1.				
KW	Hydrolase; Nucleotide metabolism.				
CONFLICT	156 156				
FT	SEQUENCE 178 AA; 19219 MW; 477C885A94AD16F1 CRC64;				
S -> F (IN REF. 1; AAB26434).					
477C885A94AD16F1 CRC64;					

EMBL: Z17216; CAAT78921.1; -					
DR	EMBL: S61107; AAB26434.1; -				
DR	EMBL: U46933; AAC54895.1; -				
DR	HSSP: P06968; IDUD.				
DR	InterPro: IPRO01428; dutpase.				
DR	Pfam: PF00692; dutpase; 1.				
KW	Hydrolase; Nucleotide metabolism.				
CONFLICT	156 156				
FT	SEQUENCE 178 AA; 19219 MW; 477C885A94AD16F1 CRC64;				
S -> F (IN REF. 1; AAB26434).					
477C885A94AD16F1 CRC64;					

CC
DR
EMBL; U32176; AA93075.1; -;
DR
EMBL; AB08266; BABI0272.1; --
DR
Mendel; 7192; ARATH.PSAN.1; -
KW
Photosystem I; Photosynthesis; Transit peptide;
KL
Thylakoid; Membrane.
KW
TRANSIT 1 86 CHLOROPLAST (BY SIMILARITY).
FT
CHAIN 87 171 PHOTOSYSTEM I REACTION CENTRE SUBUNIT N.
SQ SEQUENCE 171 AA; 18429 MW; 82FE2AFF7C599C07 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 171;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 6; Conservative 0; Mismatches 0; Gaps 0;

```

      QY       7 DYRGNY 12
                |||||
      DB     100 DYRGNY 105
    
```

RESULT 11
IM23-SCHMA

IM23_SCHMA STANDARD; PRT; 218 AA.
AC P19331; 026585;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE 23 KDA INTEGRAL MEMBRANE PROTEIN (SM23).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminths; Turbellarian Platyhelminths;
OC Rhabditophora; Eulicithophora; Revertospermata; Mediofusata;
OC Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
OC Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90217533; PubMed=2324498;
RA Wright M.D., Henkle K.J., Mitchell G.F.;
RT "An immunogenic Mr 23,000 integral membrane protein of Schistosoma
mansoni worms that closely resembles a human tumor-associated
antigen.";
RL J. Immunol. 144:3195-3200(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI;
RX MEDLINE=95121434; PubMed=7821405;
RA Lee K.W., Shalaby K.A., Medhat A.M., Shi H., Yang Q., Karim A.M.,
RA Loverde P.T.;
RT "Schistosoma mansoni: characterization of the gene encoding Sm23, an
integral membrane protein.";
RL Exp. Parasitol. 80:153-158(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE TETRASPANIN (TM4SF) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M34453; AAA29900.1; -;
DR EMBL; L34755; AAA73525.1; -;
DR PIR; A43522; A43522.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
DR PROSITE; PS00421; TM4_1; 1.
KW Glycoprotein; Antigen; Transmembrane.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 13 36 POTENTIAL.
FT DOMAIN 37 55 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 56 71 POTENTIAL.
FT DOMAIN 72 82 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 83 108 POTENTIAL.
FT DOMAIN 109 183 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 184 205 POTENTIAL.
FT DOMAIN 206 218 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 151 151 D -> N (IN REF. 2).
SQ SEQUENCE 218 AA; 23684 MW; 1B31BBD34C218890 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRGV 12
Db 152 DYRGV 157
|||||
RESULT 12
DAPB_MYCBO

DAPB_MYCBO STANDARD; PRT; 271 AA.
AC P46829;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIHYDRODIPICOLINATE REDUCTASE (EC 1.3.1.26) (DHPR).
GN DAPB.
OS Mycobacterium bovis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BCG / PASTEUR;
RX MEDLINE=94292473; PubMed=8021227;
RA Cirillo J.D., Weisbrod T.R., Banerjee A., Bloom B.R., Jacobs W.R. Jr.;
RT "Genetic determination of the meso-diaminopimelate biosynthetic
pathway of mycobacteria.";
RL J. Bacteriol. 176:4424-4429(1994).
CC -1- CATALYTIC ACTIVITY: 2,3,4,5-TETRAHYDRODIPICOLINATE + NAD(P)(+) =
2,3-DIHYDRODIPICOLINATE + NAD(P)H.
CC -1- PATHWAY: BIOSYNTHESIS OF DIAMINOPIMELATE AND LYSINE FROM ASPARTATE
SEMIALDEHYDE; SECOND STEP.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DIHYDRODIPICOLINATE REDUCTASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; L24366; AAA21562.1; -;
DR InterPro; IPR000846; DapB.
DR Pfam; PF01113; DapB; 1.
DR ProDom; PD004105; DapB; 1.
DR PROSITE; PS01298; DAPB; 1.
KW Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
KW NADP.
SQ SEQUENCE 271 AA; 27679 MW; 16DE4B090E481F6F CRC64;

Query Match 42.9%; Score 6; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDE 6
Db 19 AGVIDE 24
|||||
RESULT 13
YHDF_BACSU STANDARD; PRT; 289 AA.
ID YHDF_BACSU
AC O07575;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN CITA-SSPB INTERGENIC REGION
DE (EC 1.1.1.1).
GN YHDF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES

(SDR) FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

 DR EMBL; Y14082; CAA74490.1; -;
 DR EMBL; Z99109; CAB12784.1; -;
 DR HSSP; Q12634; 1YBV.
 DR Subtilist; BG13012; yndf.
 DR InterPro; IPR002198; ADH_short.
 DR InterPro; IPR002347; Adh_short_C2.
 DR Pfam; PF00106; adh_short_1.
 DR Pfam; PF00678; adh_short_C2; 1.
 DR PROSITE; PS00061; ADH_SHORT; 1.
 DR Hypothetical protein; Oxidoreductase; Complete proteome.
 FT NP_BIND 49 73 NAD OR NADP (BY SIMILARITY).
 FT ACT_SITE 194 194 BY SIMILARITY.
 SQ SEQUENCE 289 AA; 31509 MW; 4EC96C62B9241BC1 CRC64;

 Query Match 42.9%; Score 6; DB 1; Length 289;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 5 DEDYRG 10
 Db 34 DEDYRG 39

 RESULT 14
 SERA_ARATH STANDARD; PRT; 624 AA.
 AC 004130;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE D-3-PHOSPHOGLYCERATE DEHYDROGENASE PRECURSOR (EC 1.1.1.95) (PGDH).
 GN ATG17740 OR FL1A6.8.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99085035; PubMed=9867856;
 RA Ho C.-L., Noji M., Saito M., Saito K.;
 RT "Regulation of serine biosynthesis in Arabidopsis. Crucial role of
 RT plastidic 3-phosphoglycerate dehydrogenase in non-photosynthetic
 RT tissues".
 RL J. Biol. Chem. 274:397-402(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen A., Cheuk R.F., Chin C.W., Dewar K.,
 RA Chung M.K., Con L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.F.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.B., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
 RA Millschetter J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana".
 RL Nature 408:816-820(2000).
 CC -1- CATALYTIC ACTIVITY: 3-PHOSPHOGLYCERATE + NAD(+) =
 CC 3-PHOSPHOHYDROXYPYRUVATE + NADH.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
 CC DEHYDROGENASES FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

 DR EMBL; AB003280; BAA20405.1; -;
 DR EMBL; AB010407; BAA24440.1; -;
 DR EMBL; AC034257; AAF99816.1; -;
 DR HSSP; P08326; 1FSD.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR002162; D_2_hydroxyacid_DH.
 DR Pfam; PF00369; 2-Hacid_DH; 1.
 DR Pfam; PF01842; ACT; 1.
 DR PROSITE; PS00065; D_2-HYDROXYACID_DH_1; 1.
 DR PROSITE; PS00670; D_2-HYDROXYACID_DH_2; 1.
 DR PROSITE; PS00671; D_2-HYDROXYACID_DH_3; 1.
 DR Serine biosynthesis; Oxidoreductase; NAD; Chloroplast;
 KW Transit peptide.
 FT TRANSIT 1
 FT CHAIN 1 ? CHLOROPLAST (POTENTIAL).
 FT ACT_SITE 312 312 D-3-PHOSPHOGLYCERATE DEHYDROGENASE.
 FT ACT_SITE 341 341 SUBSTRATE-BINDING (BY SIMILARITY).
 FT ACT_SITE 360 360 BY SIMILARITY.
 SQ SEQUENCE 624 AA; 66453 MW; 724370F870DEA310 CRC64;

 Query Match 42.9%; Score 6; DB 1; Length 624;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 GVIDED 7
 Db 314 GVIDED 319

 RESULT 15
 RRPL_IAFOM STANDARD; PRT; 757 AA.
 AC Q82571;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE RNA-DIRECTED RNA POLYMERASE SUBUNIT P1 (EC 2.7.7.48) (POLYMERASE BASIC
 DE 1 PROTEIN) (PBI).
 OS Influenza A virus (strain A/Fort Monmouth/1/47).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;
 OX NCBI_TaxID=11380;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97033391; PubMed=8879138;
 RA Smeek C.A., Wright K.E., Burns B.F., Thaker A.J., Brown E.G.;
 RT "Mutations in the hemagglutinin and matrix genes of a virulent
 RT Influenza virus variant, A/FW/1/47-MA, control different stages in
 RT pathogenesis".
 RL Virus Res. 44:79-95(1996).
 CC -1- SUBUNIT: INFLUENZA RNA POLYMERASE IS COMPOSED OF THREE SUBUNITS:

CC P1 (OR PB1), P2 (OR PA), AND P3 (OR PB2).
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA VIRUSES POLYMERASE PB2
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X99037; CAA67498.1; -;
 DR InterPro; IPR001407; Flu_PB1.
 DR Pfam; PF00602; Flu_PB1; 1.
 KW Transferase: RNA-directed RNA polymerase.
 SQ SEQUENCE 757 AA; 86671 MW; 7E45D6DE764A7534 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 757;
 Best Local Similarity 100.0%; Pred No. 29;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 5 DEDYRG 10
 Db 617 DEDYRG 622
 |||||

Search completed: January 31, 2002, 13:39:25
 Job time: 85 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:23 ; Search time 46.78 seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-79
Perfect score: 14
Sequence: 1 AGVVDYDYGKVK 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	147	1 DUT_YEAST	P3317 saccharomyc
2	6	42.9	123	1 ATPE_HELPJ	Q9282 helicobacte
3	6	42.9	123	1 ATPE_HELPJ	P56084 helicobacte
4	6	42.9	132	1 MERR_BACSR	P22853 bacillus sp
5	6	42.9	185	1 VG16_HAEIN	P71387 haemophilus
6	6	42.9	316	1 TALB_ECOLI	P30148 escherichia
7	6	42.9	491	1 TRPE_NEIGO	Q9w000 neisseria g
8	6	42.9	491	1 TRPE_NEIMA	Q9xaz0 neisseria m
9	6	42.9	491	1 TRPE_NEIMB	P56995 neisseria m
10	6	42.9	491	1 TRPE_NEIMC	Q9s358 neisseria m
11	6	42.9	1434	1 RPOB_UREPA	Q9p9v6 ureaplasma
12	5	35.7	67	1 Y737_ARCFU	Q29521 archaeoglob
13	5	35.7	72	1 VG18_BPMU	Q38625 bacterioph
14	5	35.7	98	1 CMGC_BACSU	P25955 bacillus su
15	5	35.7	110	1 CU16_MANSE	Q25504 manduca sex
16	5	35.7	111	1 KV31L_MOUSE	P01664 mus musculu
17	5	35.7	113	1 PURL_SNP7	Q55041 synchococc
18	5	35.7	114	1 DCHS_LACBU	P04193 lactobacill
19	5	35.7	124	1 RS12_TREPA	O83271 treponema p
20	5	35.7	128	1 LRP_ORYSA	Q03200 oryza sativ
21	5	35.7	129	1 ATPZ_VIBAL	P12983 vibrio algi
22	5	35.7	141	1 AMPM_MYCCA	Q59509 mycoplasma
23	5	35.7	143	1 LGB1_VICFA	P02232 vicia faba
24	5	35.7	149	1 RL19_ARCFU	O28372 archaeoglob
25	5	35.7	150	1 DUT_AQUAE	O66592 aquifex aeo
26	5	35.7	158	1 SODC_ONCVO	P24706 onchocerca
27	5	35.7	159	1 DUT_CANAL	P43058 candida alb
28	5	35.7	162	1 Y269_AQUAE	O66626 aquifex aeo
29	5	35.7	167	1 NUGM_MACRO	P52670 macropus ro
30	5	35.7	168	1 YNP7_YEAST	P53897 saccharomyc
31	5	35.7	175	1 RIMM_PSEAE	Q9hxo0 pseudomonas
32	5	35.7	179	1 YCXE_CVAPA	P48335 cyanophora
33	5	35.7	181	1 NDKM_COLLI	P87355 columba liv

34	5	35.7	186	1 CHS2_USTMA	P30599 ustilago ma
35	5	35.7	192	1 EPO_MOUSE	P07321 mus musculu
36	5	35.7	193	1 HIS7_SULSO	O33773 sulfolobus
37	5	35.7	202	1 VMAT_RABVN	P25224 rabies viru
38	5	35.7	214	1 VCOL_VARV	P33659 variola vir
39	5	35.7	221	1 FIXW_RHILE	P14312 rhizobium l
40	5	35.7	226	1 TPIS_RHIET	P96985 rhizobium e
41	5	35.7	228	1 C79B_MOUSE	P15530 mus musculu
42	5	35.7	229	1 C79B_HUMAN	P40259 homo sapien
43	5	35.7	234	1 Y352_THEMA	Q9wy17 thermotoga
44	5	35.7	238	1 Y206_CHLMU	Q9plai chlamydia m
45	5	35.7	239	1 RNL2_LENED	P81296 lentinula e

ALIGNMENTS

RESULT 1				
DUT_YEAST				
ID	DUT_YEAST	STANDARD;	PRT;	147 AA.
AC	P33317;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	DEOXYRIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)			
DE	(DUTPASE) (DUTP PYROPHOSPHATASE).			
GN	DUT1 OR YBR252W OR YBR1705.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94038925; PubMed=8223452;			
RA	Gadsden M.H., McIntosh E.M., Game J.C., Wilson P.J., Haynes R.H.;			
RT	"dUTP pyrophosphatase is an essential enzyme in Saccharomycetes cerevisiae."			
RT	cerevisiae."			
RL	EMBO J. 12:4425-4431(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C;			
RX	MEDLINE=94078675; PubMed=8256522;			
RA	Doignon F., Biteau N., Aigle M., Crouzet M.;			
RT	"The complete sequence of a 6794 bp segment located on the right arm of chromosome II of Saccharomycetes cerevisiae. Finding of a putative DUTPase in a yeast."			
RT	Yeast 9:1131-1137(1993).			
RL				
CC	-!- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT URACIL CANNOT BE INCORPORATED INTO DNA.			
CC	-!- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.			
CC	-!- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.			
CC	-!- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC				
DR	EMBL; X74263; CAAS2322.1; -			
DR	EMBL; L20296; AAA65611.1; -			
DR	EMBL; Z36121; CAAS5215.1; -			
DR	PIR; S38189; S38189.			
DR	SGD; S0000456; DUT1.			
DR	InterPro; IPR001428; DUTPase.			
DR	Fram; PF00692; DUTPase; 1.			
KW	Hydrolase; Nucleotide metabolism.			
FT	CONFLICT 10 10 N-> K (IN REF. 2).			
SQ	SEQUENCE 147 AA; 15293 MW; 6F1E87A692A061F6 CRC64;			

Query Match 100.0%; Score 14; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 6e-08; 0; Mismatches 0; Indels 0; Gaps 0;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGVVDRTYTGVEVK 14
 DB 81 AGVVDRTYTGVEVK 94

RESULT 2

ATPE_HELPJ STANDARD; PRT; 123 AA.
 AC Q9ZK82;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).
 GN ATPC OR JHP1059.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Meir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 gastric pathogen Helicobacter pylori";
 RL Nature 397:176-180(1999).
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 CC GRADIENT ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001533; AB006638.1;
 CC HSP; P00832; IBSH.
 CC InterPro: IPR001469; ATP-synt_DE.
 CC Pfam: PF00401; ATP-synt_DE; 1.
 CC ProDom: PD000944; ATP-synt_DE; 1.
 CC HydroLase: ATP synthesis; CF(1); Hydrogen ion transport;
 CC Complete proteome.
 CC SEQUENCE 123 AA; 13271 MW; 19F561ED5B849761 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 123;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 YTGEVK 13
 DB 16 YTGEVK 21

RESULT 3

ATPE_HELPJ STANDARD; PRT; 123 AA.
 ID MERR_BACSR

P56084;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).
 GN ATPC OR HPI131.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=26695 / ATCC 700392;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 pylori";
 RL Nature 388:539-547(1997).
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 CC GRADIENT ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000619; AAD08173.1; ALT_INIT.
 CC HSP; P00832; IAQT.
 CC TIGR; HP1131;
 CC InterPro: IPR001469; ATP-synt_DE.
 CC Pfam: PF00401; ATP-synt_DE; 1.
 CC ProDom: PD000944; ATP-synt_DE; 1.
 CC HydroLase: ATP synthesis; CF(1); Hydrogen ion transport;
 CC Complete proteome.
 CC SEQUENCE 123 AA; 13357 MW; 19EDDA4D2B3A5461 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 123;
 Best Local Similarity 100.0%; Pred. No. 8.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 YTGEVK 13
 DB 16 YTGEVK 21

RESULT 4

MERR_BACSR STANDARD; PRT; 132 AA.
 ID MERR_BACSR
 AC P22853;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-OCT-1993 (Rel. 27, Last annotation update)
 DE MERCURIC RESISTANCE OPERON REGULATORY PROTEIN.
 GN MERR.
 OS Bacillus sp. (strain RC607).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1420;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89123092; PubMed=2536669;
RA Wang Y., Moore M., Levinson H.S., Silver S., Walsh C., Mahler I.;
RT "Nucleotide sequence of a chromosomal mercury resistance determinant
from a Bacillus sp. with broad-spectrum mercury resistance.";
RL J. Bacteriol. 171:83-92(1989).
RN [2]
RP FUNCTION
RX MEDLINE=89123021; PubMed=2492496;
RA Helmann J.D., Wang Y., Mahler I., Walsh C.T.;
RT "Homologous metalloregulatory proteins from both gram-positive and
gram-negative bacteria control transcription of mercury resistance
operons";
RL J. Bacteriol. 171:222-229(1989).
RN [3]
RP MUTAGENESIS.
RX MEDLINE=90161989; PubMed=2305262;
RA Helmann J.D., Ballard B.T., Walsh C.T.;
RT "The MerR metalloregulatory protein binds mercuric ion as a
tricoordinate, metal-bridged dimer";
RL Science 247:946-948(1990).
RN
CC -1- FUNCTION: MEDIATES THE MERCURY-DEPENDENT INDUCTION OF MERCURY
RESISTANCE OPERON. IN THE ABSENCE OF MERCURY MERR REPRESSSES
TRANSCRIPTION BY BINDING TIGHTLY TO THE MERR OPERATOR REGION;
WHEN MERCURY IS PRESENT THE DIMERIC COMPLEX BINDS A SINGLE ION
AND BECOMES A POTENT TRANSCRIPTIONAL ACTIVATOR, WHILE REMAINING
BOUND TO THE MERR SITE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; AF138877; AAA83973.1; -
DR PIR; A32227; A32227.
DR PIR; A32239; A32239.
DR InterPro; IPR000551; HTH_MerR.
DR Pfam; PF00376; merK; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW Transcription regulation; Activator; Repressor; Mercuric resistance;
KW Mercury; DNA-binding.
FT DNA_BIND 5 24 H-T-H MOTIF (POTENTIAL).
FT METAL 79 79 HG(2+).
FT METAL 114 114 HG(2+).
FT METAL 123 123 HG(2+).
FT METAL 123 123 HG(2+).
FT MUTAGEN 79 79 C->A,H: LOSS OF HG BINDING.
FT MUTAGEN 114 114 C->A: LOSS OF HG BINDING.
FT MUTAGEN 123 123 C->A,H: LOSS OF HG BINDING.
SQ SEQUENCE 132 AA; 15971 MW; 6557FBF1FB95B635 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVVDRO 7
Db 70 GVVDRO 75
RESULT 5
VG16_HAEIN

VG16_HAEIN STANDARD; PRT; 185 AA.
P71387;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MU-LIKE PROPHAGE FLUO MU PROTEIN GP16.
GN H11488.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD.";
RL Science 269:496-512(1995).
CC -1- SIMILARITY: STRONG, TO PHAGE MU PROTEIN GP16.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; U32826; AAC23133.1; -
DR TIGR; H11488; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 185 AA; 20954 MW; 23C80042B2F22449 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGEVKV 14
Db 77 TGEVKV 82
RESULT 6
TALB_ECOLI STANDARD; PRT; 316 AA.
AC P30148;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSALDOLASE B (EC 2.2.1.2).
GN TALB OR B0008.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96011394; PubMed=7592346;
RA Sprenger G.A., Schorken U., Sprenger G., Sahm H.;
RT "Transaldolase B of Escherichia coli K-12: cloning of its gene, talB,
and characterization of the enzyme from recombinant strains.";
RL J. Bacteriol. 177:5930-5936(1995).
RN [2]

RC SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA Iida A., Teshiba S., Mizobuchi K.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RL "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RN Nucleic Acids Res. 20:3305-3308(1992).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12";
RT Science 277:1453-1474(1997).
[5]
RN SEQUENCE OF 1-11.
RC STRAIN-K12 / W3110;
RX Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Baillou A.,
RA Hochstrasser D.F.;
RL Submitted (SEP-1994) to the SWISS-PROT data bank.
[6]
RN SEQUENCE OF 1-12.
RC STRAIN-K12 / EMC2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RL "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RN Electrophoresis 18:1239-1313(1997).
[7]
RN PRESENCE OF TWO TRANSALDOLASES IN E. COLI.
RC Sprenger G.A.;
RL Unpublished observations (JUN-1993).
[8]
RN X-RAY CRYSTALLOGRAPHY (1.87 ANGSTROMS).
RC MEDLINE=96399717; PubMed=880555;
RA Jia J., Huang W., Schoerlen U., Sham H., Sprenger G.A., Lindqvist Y.,
RA Schneider G.;
RL "Crystal structure of transaldolase B from Escherichia coli suggests
RT a dimeric structure of the alpha/beta barrel within the class I
RL aldolase family.";
RN Structure 4:715-724(1996).
[9]
RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RC MEDLINE=97160483; PubMed=9007983;
RA Jia J., Schorken U., Lindqvist Y., Sprenger G.A., Schneider G.;
RL "Crystal structure of the reduced Schiff-base intermediate complex of
RT transaldolase B from Escherichia coli: mechanistic implications for
RT class I aldolases.";
RN Protein Sci. 6:119-124(1997).
[10]
RN FUNCTION: TRANSALDOLASE IS IMPORTANT FOR THE BALANCE OF
RC METABOLITES IN THE PENTOSE-PHOSPHATE PATHWAY.
CC METABOLITES IN THE PENTOSE-PHOSPHATE PATHWAY.
CC -1- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-GLYCERALDEHYDE
CC 3-PHOSPHATE = D-ERYTHROSE 4-PHOSPHATE + D-FRUCTOSE 6-PHOSPHATE.
CC -1- PATHWAY: HOMODIMER.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO THE TRANSALDOLASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC or send an email to license@isb-sib.ch.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: S80045; AAB47022.1; -;
CC EMBL: D13161; BAA21822.1; -;
CC EMBL: D10483; BAA01290.1; -;
CC EMBL: AE000111; AAC73119.1; -;
CC DR PIR: S40535; S40535.
CC DR PDB: 1ONR; 12-MAR-97.
CC DR PDB: 1UCW; 07-JUL-97.
CC DR SWISS-2DPAGE: P30148; COLI.
CC EcoGene: EG11556; talB.
CC InterPro: IPR001585; Transaldolase.
CC Pfam: PF00923; Transaldolase; 1.
CC PROSITE: PS00958; TRANSALDOLASE-2; 1.
CC PROSITE: PS01054; TRANSALDOLASE-1; 1.
CC Transferrase: Pentose shunt; Multigene family; 3D-structure;
CC Complete proteome.
CC INIT_MET 0
CC ACT_SITE 131
CC FT ACT_SITE 131
CC SQ SEQUENCE 316 AA; 35088 MW; 2D71C44DFCB55523 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 316;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YTGEVK 13
Db 264 YTGEVK 269
| | | | |
| | | | |

RESULT 7
TRPE NEIGO
ID TRPE NEIGO STANDARD; PRT; 491 AA.
AC 35WQ0; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN TRPE
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 70825 / FA 1090, AND MS11;
RX MEDLINE=99348393; PubMed=10417653;
RA Zhu P., Morelli G., Achtman M.;
RL "The opa and (psi)opch regions in Neisseria: genes, pseudogenes,
RT deletions, insertion elements and DNA islands.";
RN Mol. Microbiol. 33:635-650(1999).
CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
CC PYRUVATE + L-GLUTAMATE.
CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SUBUNIT: Tetramer of two components I and two components II (by
CC similarity).
CC -1- MISCELLANEOUS: Component I catalyzes the formation of anthranilate
CC using ammonia rather than glutamine, whereas component II provides
CC glutamine amidotransferase activity.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ242839; CAB45014.1; -;
CC EMBL: AJ242840; CAB44949.1; -;
CC

DR InterPro: IPR000350; Chorismate_bind.
 DR Pfam: PF00425; chorismate_bind; 1.
 DR PRINTS; PR00095; ANTSNTHASEI.
 DR ProDom: PD000779; Chorismate_bind; 1.
 KW Tryptophan biosynthesis; Lyase.
 SQ SEQUENCE 491 AA; 54748 MW; A810F7B2304FE47F CRC64;

Query Match 42.9%; Score 6; DB 1; Length 491;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TGEVKV 14
 Db 351 TGEVKV 356
 |||||

RESULT 8
 TRPE_NEIMA STANDARD; PRT; 491 AA.
 AC Q9XAZ0;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
 GN TRPE OR NMA1247.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies S.R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 Jajels K., Leather S., Moule S., Mungall K., Quail M.A.,
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491.";
 RL Nature 404:502-506(2000).
 CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
 PYRUVATE + L-GLUTAMATE.
 CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
 CC -1- SUBUNIT: Tetramer of two components I and two components II (by
 similarity).
 CC -1- MISCELLANEOUS: Component I catalyzes the formation of anthranilate
 using ammonia rather than glutamine, whereas component II provides
 glutamine amidotransferase activity.
 CC -1- SIMILARITY: Belongs to the anthranilate synthase component I
 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ242841; CAB44973.1; -
 DR EMBL: AL162755; CAB84502.1; -
 DR InterPro: IPR000350; Chorismate_bind.
 DR Pfam: PF00425; chorismate_bind; 1.

DR PRINTS; PR00095; ANTSNTHASEI.
 DR ProDom: PD000779; Chorismate_bind; 1.
 KW Tryptophan biosynthesis; Lyase; Complete proteome.
 SQ SEQUENCE 491 AA; 54722 MW; 035A92E12707B660 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 491;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 TGEVKV 14
 Db 351 TGEVKV 356
 |||||

RESULT 9
 TRPE_NEIMB STANDARD; PRT; 491 AA.
 AC P56995;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
 GN TRPE OR NMB1021.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eisen J.A., Ketchum K.A., Hood D.W., Bदन J.F., Dodson R.J.,
 Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
 Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 Gill J.H., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
 Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58.";
 RL Science 287:1809-1815(2000).
 CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE = ANTHRANILATE +
 PYRUVATE + L-GLUTAMATE.
 CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
 CC -1- SUBUNIT: Tetramer of two components I and two components II (by
 similarity).
 CC -1- MISCELLANEOUS: Component I catalyzes the formation of anthranilate
 using ammonia rather than glutamine, whereas component II provides
 glutamine amidotransferase activity.
 CC -1- SIMILARITY: Belongs to the anthranilate synthase component I
 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE002452; AAF41421.1; -
 DR TIGR: NMB1021; -
 DR InterPro: IPR000350; Chorismate_bind.
 DR Pfam: PF00425; Chorismate_bind; 1.
 DR ProDom: PD000779; Chorismate_bind; 1.
 KW Tryptophan biosynthesis; Lyase; Complete proteome.
 SQ SEQUENCE 491 AA; 54702 MW; 815236B1334D122C CRC64;

Query Match 42.9%; Score 6; DB 1; Length 491;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 9 TGEVKV 14
Db 351 TGEVKV 356

RESULT 10
TRPE_NEIMC STANDARD; PRT; 491 AA.
AC Q9S358;
DT 30-MAY-2000 (Rel. 39, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ANTHRANILATE SYNTHASE COMPONENT I (EC 4.1.3.27).
GN TRPE.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=135720;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 700532 / FAM18 / SEROGROUP C;
RX MEDLINE=99348393; PubMed=10417653;
RA Zhu P., Morelli G., Achtman M.;
RT "The opa and pseudo opa regions in Neisseria: genes, pseudogenes, deletions, insertion elements and DNA islands.";
RL Mol. Microbiol. 33:635-650(1999).
CC -1- CATALYTIC ACTIVITY: CHORISMATE + L-GLUTAMINE - ANTHRANILATE +
CC -1- PYRUVATE + L-GLUTAMATE.
CC -1- PATHWAY: FIRST STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
CC -1- SUBUNIT: Tetramer of two components I and two components II (by similarity).
CC -1- MISCELLANEOUS: COMPONENT I CATALYZES THE FORMATION OF ANTHRANILATE USING AMMONIA RATHER THAN GLUTAMINE, WHEREAS COMPONENT II PROVIDES GLUTAMINE AMIDOTRANSFERASE ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE SYNTHASE COMPONENT I FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ242842; CAB44980.1;
CC InterPro: IPR000350; Chorismate_bind.
CC Pfam: PF00425; chorismate_bind; 1.
CC PRINTS: PR00095; ANTSNTHASEI.
CC ProDom: PD000779; Chorismate_bind; 1.
CC TrpTophan biosynthesis; Lyase.
SQ SEQUENCE 491 AA; 54719 MW; FC7053315D19EA7B CRC64;

Query Match 42.9%; Score 6; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 TGEVKV 14
Db 351 TGEVKV 356

RESULT 11
RPOB_UREPA STANDARD; PRT; 1434 AA.
AC Q9POV6;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT).
GN RPOB OR U0187.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/clostridium group; Molllicutes;
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SERVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y., Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma urealyticum.";
RL Nature 407:757-762(2000).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE + RNA(N).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1 BETA CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE002118; AAF30594.1;
CC InterPro: IPR001572; RNA_POL_B.
CC Pfam: PF00562; RNA_POL_B; 2.
CC PROSITE: PS01166; RNA_POL_BETA; 1.
CC Transferrase; Transcription; DNA-directed RNA polymerase; Complete proteome.
SQ SEQUENCE 1434 AA; 162022 MW; 2841BFA50C19BA7A CRC64;

Query Match 42.9%; Score 6; DB 1; Length 1434;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 YTGVEVK 13
Db 106 YTGVEVK 111

RESULT 12
Y737_ARCFU STANDARD; PRT; 67 AA.
AC O29521;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOVTHEICAL PROTEIN AF0737.
GN AF0737.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D., Richardson D.L., Kervilave A.R., Graham D.E., Kyriakides N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Relch C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
```

RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE001054; AAB90514.1; -
 DR TIGR; AF0737; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 67 AA; 7296 MW; 6F103AABBBE56630 CRC64;

 Query Match 35.7%; Score 5; DB 1; Length 67;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 GEVKV 14
 Db 56 GEVKV 60
 I I I I I

 RESULT 13
 ID VG18_BPMU STANDARD; PRT; 72 AA.
 AC Q38625;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN GP18 (E18 PROTEIN).
 GN 18 OR E18.
 OS Bacteriophage Mu.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
 OX NCBI_TaxID=10677;
 RN [1]
 RA Stoddard S.F., Howe M.M.;
 RP Submitted (SEP-1987) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RA Pries H., Brauer B., Schmidt C., Kamp D.;
 RP "Sequence of the left end of Mu.";
 RT (in) Symonds N., Toussaint A., van de Putte P., Howe M.M. (eds.);
 RL Phage Mu, pp.277-296, Cold Spring Harbor Laboratory Press,
 RL New York (1987).
 [3]
 RN SEQUENCE FROM N.A.
 RA Morgan G., Hatfull G., Hendrix R.;
 RP "Genome of bacteriophage Mu and comparison with the Haemophilus
 RT influenzae Mu-like prophage Flumu.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y00419; CAA68476.1; -
 DR EMBL; M64097; AAA32410.1; -
 DR EMBL; AF083977; AAF01095.1; -
 KW Hypothetical protein.

SQ SEQUENCE 72 AA; 8444 MW; B29272F519B8E068 CRC64;

 Query Match 35.7%; Score 5; DB 1; Length 72;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 RDTYG 10
 Db 34 RDTYG 38
 I I I I I

 RESULT 14
 ID CMGC_BACSU STANDARD; PRT; 98 AA.
 AC P25955;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE COMG OPERON PROTEIN 3 PRECURSOR.
 GN CMGC OR COMG3.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90008773; PubMed=2507524;
 RA Albano M., Breitling R., Dubnau D.A.;
 RT "Nucleotide sequence and genetic organization of the Bacillus
 RT subtilis comG operon.";
 RL J. Bacteriol. 171:5386-5404(1989).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=168 / JH642;
 RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
 RA Sato T., Takeuchi M.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 [3]
 RN CHARACTERIZATION
 RP MEDLINE=90170868; PubMed=1968455;
 RA Breitling R., Dubnau D.A.;
 RT "A membrane protein with similarity to N-methylphenylalanine pills
 RT is essential for DNA binding by competent Bacillus subtilis.";
 RL J. Bacteriol. 172:1499-1508(1990).
 [4]
 RN FUNCTION.
 RP MEDLINE=98083053; PubMed=9422590;
 RA Chung Y.S., Dubnau D.A.;
 RT "All seven comG open reading frames are required for DNA binding
 RT during transformation of competent Bacillus subtilis.";
 RL J. Bacteriol. 180:41-45(1998).
 [5]
 RN SUBCELLULAR LOCATION AND PROBABLE DISULFIDE BOND.
 RP MEDLINE=98389321; PubMed=9723928;
 RA Chung Y.S., Breit F., Dubnau D.A.;
 RT "Cell surface localization and processing of the ComG proteins.
 RT required for DNA binding during transformation of Bacillus subtilis.";
 RL Mol. Microbiol. 29:905-913(1998).
 CC -1- FUNCTION: REQUIRED FOR TRANSFORMATION AND DNA-BINDING.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: THE UNPROCESSED FORM IS AN INTEGRAL MEMBRANE
 CC PROTEIN WITH ITS C-TERMINUS OUTSIDE THE MEMBRANE. UPON CLEAVAGE,
 CC IT IS TRANSLOCATED TO THE OUTER FACE OF THE MEMBRANE.
 CC -1- PTM: PROCESSING OF COMG IN COMPETENT CELLS REQUIRES COMC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```
CC EMBL; M29691; AAA83369.1; -
DR EMBL; D84432; BAA12535.1; -
DR EMBL; Z99116; CAA14402.1; -
DR PIR; D30338; D30338.
DR PIR; A35133; A35133.
DR Subtilisin; BG10485; COMGC.
DR InterPro; IPR001120; Prok_N_methylto.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
KW Transport; Methylation; Transmembrane; Complete proteome.
FT CHAIN 1 5 BY SIMILARITY.
FT TRANSMEM 6 98 COMG OPERON PROTEIN 3.
FT MOD_RES 6 26 POTENTIAL.
FT DISULFID 41 81 METHYLATION (BY SIMILARITY).
FT SEQUENCE 98 AA; 10850 MW; 17B8152CAFE2E4C9 CRC64;
SQ
Query Match 35.7%; Score 5; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 GEVKV 14
Db 92 GEVKV 96
|||||
RESULT 15
CUI6 MANSE
ID CUI6 MANSE STANDARD; PRT; 110 AA.
AC Q25504;
DC 12-JUL-1998 (Rel. 36, Created)
DT 13-JUL-1998 (Rel. 36, Last sequence update)
DE LARVAL CUTICLE PROTEIN 16/17 PRECURSOR.
GN LCP16/17.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Spingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Abdominal epidermis;
RX MEDLINE=89171584; PubMed=2924995;
RA Horodyski F.M., Riddiford L.M.;
RT "Expression and hormonal control of a new larval cuticular multigene
family at the onset of metamorphosis of the tobacco hornworm.";
RL Dev. Biol. 132:292-303(1989).
CC -1- FUNCTION: COMPONENT OF THE CUTICLE OF THE LARVA OF TOBACCO
HORNWORM. SEEMS TO PARTICIPATE TO THE FORMATION OF 5- TO 10-FOLD
THINNER LAMELLAE IN THE ENDOCUTICLE.
CC -1- TISSUE SPECIFICITY: SPECIFIC TO THE EPIDERMIS. EXPRESSED IN ALL
THE EPIDERMAL CELLS OF DAY 3 LARVAE EXCEPT FOR THE BRISTLE CELLS
AND THOSE AT THE MUSCLE ATTACHMENT SITES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED LATE ON THE PENULTIMATE DAY OF
FEEDING IN THE FIFTH LARVAL INSTAR. HIGHEST LEVELS THROUGHOUT THE
FINAL DAY OF FEEDING. BARELY DETECTABLE BY THE DAY OF WANDERING.
CC EXPRESSION STOPS WHEN THE METAMORPHOSIS BEGINS.
CC -1- INDUCTION: IN DAY 1 EPIDERMIS, BY EXPOSURE TO HYDROXYCDYSONE IN
VITRO WITHOUT JUVENILE HORMONE.
CC -1- SIMILARITY: CONTAINS A CUTICLE CONSENSUS DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M25486; AAA50287.1; -
DR InterPro; IPR000618; Insect_cuticle.
```

```
DR Pfam; PF00379; Insect_cuticle; 1.
DR PRINTS; PR00947; CUTICLE.
DR PROSITE; PS00233; CUTICLE; 1.
KW Structural protein; Cuticle; Signal; Polymorphism.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 110 LARVAL CUTICLE PROTEIN 16/17.
FT DOMAIN 68 71 POLY-VAL.
FT VARIANT 64 64 K -> D (COULD BE A POLYMORPHISM).
SQ SEQUENCE 110 AA; 12240 MW; 998BCFBD3318F691 CRC64;
Query Match 35.7%; Score 5; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 TGEVK 13
Db 52 TGEVK 56
|||||
Search completed: January 31, 2002, 13:39:24
Job time: 84 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:22 ; Search time 46.78 seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-78
Perfect score: 14
Sequence: 1 VGLIDSDYQGGLMI 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	14	100.0	151	DUT_ECOLI	P06968 escherichia
2	14	100.0	154	DUT_BUCAI	P57623 buchnera ap
3	10	71.4	151	DUT_HAEN	P43792 haemophilus
4	10	71.4	151	DUT_PASMU	P57914 pasteurella
5	10	71.4	152	DUT_COXBU	Q45920 coxiella bu
6	6	42.9	293	YMO2_RHIME	P49306 rhizobium m
7	6	42.9	586	HO_YEAST	P09932 saccharomyc
8	6	42.9	805	SYFB_MYCPN	P75563 mycoplasma
9	6	42.9	848	AMPN_LACLA	Q48656 lactococcus
10	6	42.9	1071	VATA_YEAST	P17255 saccharomyc
11	5	35.7	69	PAP2_SPVKA	P23334 swinepox vi
12	5	35.7	99	YGAV_ECOLI	P77295 escherichia
13	5	35.7	128	YJZ2_YEAST	P47093 saccharomyc
14	5	35.7	141	DUT_CHVPI	Q41033 paramecium
15	5	35.7	142	DUT_SPVKA	P32208 swinepox vi
16	5	35.7	143	YL23_ARCFU	Q28157 archaeeoglob
17	5	35.7	145	DUT_CHLPN	Q929C2 chlamydia p
18	5	35.7	146	DUT_ZYMMO	Q9X3X5 zymomonas m
19	5	35.7	148	DUT_LPT5	Q48500 bacterioph
20	5	35.7	148	PBPI_DROME	P54191 drosophila
21	5	35.7	160	PTP2_NVOP	Q10273 orgyia pseu
22	5	35.7	161	SP2G_BACTK	P26767 bacillus th
23	5	35.7	163	DUT_ADEG8	Q9VYS0 avian adeno
24	5	35.7	175	CRGL_XENLA	Q06254 xenopus lae
25	5	35.7	175	HSLV_BUCAI	P57115 buchnera ap
26	5	35.7	188	RECX_MYCSM	P94965 mycobacteri
27	5	35.7	199	YE74_HAEN	Q57213 haemophilus
28	5	35.7	203	MSG1_MOUSE	P97769 mus musculu
29	5	35.7	212	KTHV_SYNY3	Q55593 synecocyst
30	5	35.7	216	FGFH_HUMAN	Q60258 homo sapien
31	5	35.7	216	FGFH_MOUSE	O70627 mus musculu
32	5	35.7	220	PAP2_CAFVK	P19748 capripoxvir
33	5	35.7	241	YWBG_BACSU	P39590 bacillus su

RESULT 1

ID	DUT_ECOLI	STANDARD;	PRT;	151 AA.
AC	P06968:			
DT	01-APR-1988 (Rel. 07, Created)			
DT	01-APR-1988 (Rel. 07, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	DEOXYRIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)			
DE	(DUTPASE) (DUTP PYROPHOSPHATASE).			
GN	DUT OR DNAS OR SOF OR B3640.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12;			
RX	MEDLINE=84057777; PubMed=6139280;			
RA	Lundberg L.G., Thoreson H.-O., Karlstroem O.H., Nyman P.O.;			
RT	"Nucleotide sequence of the structural gene for dUTPase of			
RT	Escherichia coli K-12.;"			
RL	EMBO J. 2:967-971(1983).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-K12 / MG1655;			
RX	MEDLINE=93315143; PubMed=7686882;			
RA	Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;			
RT	"DNA sequence and analysis of 136 kilobases of the Escherichia coli			
RT	genome: organizational symmetry around the origin of replication.;"			
RL	Genomics 16:551-561(1993).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=92158084; PubMed=1311056;			
RA	Cedergren-Zeppezauer E.S., Larsson G., Nyman P.O., Dauter Z.,			
RA	Wilson K.S.;			
RT	"Crystal structure of a dUTPase.;"			
RL	Nature 355:740-743(1992).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).			
RX	MEDLINE=96227973; PubMed=8646539;			
RA	Larsson G., Svensson L.A., Nyman P.O.;			
RT	"Crystal structure of the Escherichia coli dUTPase in complex with a			
RT	substrate analogue (dUDP).;"			
RL	Nat. Struct. Biol. 3:532-538(1996).			
RN	[5]			
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).			
RX	MEDLINE=98437602; PubMed=9757088;			
RA	Dauter Z., Wilson K.S., Larsson G., Nyman P.O.,			
RA	Cedergren-Zeppezauer E.S.;			
RT	"The refined structure of dUTPase from Escherichia coli.;"			
RL	Acta Crystallogr. D 54:735-749(1998).			
RN	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.05 ANGSTROMS).			
RX	PubMed=11375495;			
RA	Gonzalez A., Larsson G., Persson R., Cedergren-Zeppezauer E.S.;			
RT	"Atomic resolution structure of Escherichia coli dUTPase determined ab			

P42360 streptococc
P42729 rhizobium l
Q91755 xenopus lae
P55607 rhizobium s
Q03417 zymomonas m
O26230 methanobact
Q92mv6 helicobacte
P56080 helicobacte
P13801 bacillus su
P57630 buchnera ap
Q94660 plasmodium
Q9zchl rickettsia

RT initio.
 RL Acta Crystallogr. D 57:767-774 (2001).
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 CC URACIL CANNOT BE INCORPORATED INTO DNA.
 CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X01714; CAA25859.1; -
 CC EMBL; V01578; CAA24897.1; -
 CC EMBL; L10328; AAA61993.1; -
 CC EMBL; AE00441; AAC76664.1; -
 CC PIR; Q00497; WPECDD.
 CC PDB; 1DUP; 01-SEP-95.
 CC PDB; 1DUD; 08-NOV-96.
 CC PDB; 1EUV; 03-MAY-00.
 CC SWISS-2DPAGE; P06968; COLI.
 CC ECO2DBASE; C017.2; 6TH EDITION.
 CC EcoGene; EG10251; dut.
 CC InterPro; IPR001428; dUTPase.
 CC Pfam; PF00692; dUTPase; 1.
 CC Hydrolase; Nucleotide metabolism; 3D-structure; Complete proteome.
 KW SEQUENCE 151 AA; 16155 MW; 98FA3DE0BC70FFB2 CRC64;
 SQ
 Query Match 100.0%; Score 14; DB 1; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.3e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VGLIDSDYQGOLMI 14
 Db 85 VGLIDSDYQGOLMI 98
 RESULT 2
 ID DUT_BUCAI STANDARD; PRT; 154 AA.
 AC P57623;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DEXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDHYDROLASE (EC 3.6.1.23)
 DE (DUTPASE) (DUTP PYROPHOSPHATASE).
 GN DUT OR BU560.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TOKYO 1998;
 RX MEDLINE-20445173; PubMed-10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT Buchnera sp. APS."
 RL Nature 407:81-86 (2000).
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 CC URACIL CANNOT BE INCORPORATED INTO DNA (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X01714; CAA25859.1; -
 CC EMBL; V01578; CAA24897.1; -
 CC EMBL; L10328; AAA61993.1; -
 CC EMBL; AE00441; AAC76664.1; -
 CC PIR; Q00497; WPECDD.
 CC PDB; 1DUP; 01-SEP-95.
 CC PDB; 1DUD; 08-NOV-96.
 CC PDB; 1EUV; 03-MAY-00.
 CC SWISS-2DPAGE; P06968; COLI.
 CC ECO2DBASE; C017.2; 6TH EDITION.
 CC EcoGene; EG10251; dut.
 CC InterPro; IPR001428; dUTPase.
 CC Pfam; PF00692; dUTPase; 1.
 CC Hydrolase; Nucleotide metabolism; 3D-structure; Complete proteome.
 KW SEQUENCE 151 AA; 16155 MW; 98FA3DE0BC70FFB2 CRC64;
 SQ

CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; APO01119; BAB13250.1; -
 CC InterPro; IPR001428; dUTPase.
 CC Pfam; PF00692; dUTPase; 1.
 KW Hydrolase; Nucleotide metabolism; Complete proteome.
 SQ SEQUENCE 154 AA; 16936 MW; E5B5FDEBFB09E920 CRC64;
 Query Match 100.0%; Score 14; DB 1; Length 154;
 Best Local Similarity 100.0%; Pred. No. 1.3e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VGLIDSDYQGOLMI 14
 Db 85 VGLIDSDYQGOLMI 98
 RESULT 3
 ID DUT_HAEIN STANDARD; PRT; 151 AA.
 AC P43792;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DEXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDHYDROLASE (EC 3.6.1.23)
 DE (DUTPASE) (DUTP PYROPHOSPHATASE).
 GN DUT OR H10954.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20 / ATCC 51907;
 RX MEDLINE-95350630; PubMed-7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd."
 RL Science 269:496-512 (1995).
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 CC URACIL CANNOT BE INCORPORATED INTO DNA (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
 CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U32776; AAC22615.1; -

```
DR HSP; P06968; 1DUD.
DR TIGR; H10954; -.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR Deoxyuridine 5'-triphosphate metabolism; Complete proteome.
SQ SEQUENCE 151 AA; 16445 MW; 78A45C50518FCDE2 CRC64;

Query Match 71.4%; Score 10; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGLIDSDYQG 10
Db 85 VGLIDSDYQG 94

RESULT 4
DUT_PASMU
ID DUT_PASMU STANDARD; PRT; 151 AA.
AC P57914;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPASE) (DUTP PYROPHOSPHATASE).
GN DUT OR PM154.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.D., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
CC URACIL CANNOT BE INCORPORATED INTO DNA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X79075; CAA55678.1; -.
CC HSP; P06968; 1DUD.
CC InterPro; IPR001428; dUTPase.
CC Pfam; PF00692; dUTPase; 1.
CC Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 152 AA; 16213 MW; 00041727C1882C57 CRC64;

Query Match 71.4%; Score 10; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGLIDSDYQG 10
Db 86 VGLIDSDYQG 95

RESULT 6
YMO2_RHIME
ID YMO2_RHIME STANDARD; PRT; 293 AA.
AC P49306;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 32.3 KDA PROTEIN IN MOCC-MOCA INTERGENIC REGION (ORF293).
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L5-30;
RX MEDLINE=95147842; PubMed=7845353;
RA Rosbach S., Kulpa D.A., Rosbach U., de Bruijn F.J.;
RT "Molecular and genetic characterization of the rhizopine catabolism
RL (mocABK) genes of Rhizobium meliloti L5-30.";
RL Mol. Gen. Genet. 245:11-24(1994).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
ID DUT_COXBU STANDARD; PRT; 152 AA.
AC Q45920;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPASE) (DUTP PYROPHOSPHATASE).
GN DUT.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Coxiella group; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NINE MILE PHASE I;
RA Thiele D., Willemis H., Oswald W., Krauss H.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
CC URACIL CANNOT BE INCORPORATED INTO DNA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X79075; CAA55678.1; -.
CC HSP; P06968; 1DUD.
CC InterPro; IPR001428; dUTPase.
CC Pfam; PF00692; dUTPase; 1.
CC Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 152 AA; 16213 MW; 00041727C1882C57 CRC64;

Query Match 71.4%; Score 10; DB 1; Length 152;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGLIDSDYQG 10
Db 86 VGLIDSDYQG 95

RESULT 6
YMO2_RHIME
ID YMO2_RHIME STANDARD; PRT; 293 AA.
AC P49306;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 32.3 KDA PROTEIN IN MOCC-MOCA INTERGENIC REGION (ORF293).
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L5-30;
RX MEDLINE=95147842; PubMed=7845353;
RA Rosbach S., Kulpa D.A., Rosbach U., de Bruijn F.J.;
RT "Molecular and genetic characterization of the rhizopine catabolism
RL (mocABK) genes of Rhizobium meliloti L5-30.";
RL Mol. Gen. Genet. 245:11-24(1994).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X78503; CAA55268.1; -;
CC HYPOTHETICAL PROTEIN.
CC SEQUENCE 293 AA; 32308 MW; 2D4A662356EAA2D9 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 293;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLDSD 7
Db 96 GLDSD 101

RESULT 7
ID HO_YEAST STANDARD; PRT; 586 AA.
AC P09332; Q12183;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HOMOTHALLIC SWITCHING ENDONUCLEASE.
GN HO OR YDL227C
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=67089786; PubMed=3025649;
RA Russell D.W., Jensen R., Zoller M.J., Burke J., Errede B., Smith M.,
RA Herskowitz I.,
RT Structure of the Saccharomyces cerevisiae HO gene and analysis of
RT its upstream regulatory region.;
RL Mol. Cell. Biol. 6:4281-4294 (1986).
RN [2]
RP SEQUENCE FROM N.A.
NC STRAIN=S288C;
RX MEDLINE=96120866; PubMed=8590483;
RA Meltron H., Nahon E., Ravesh D.;
RT Identification of the heterothallic mutation in HO-endonuclease of
RT S. cerevisiae using HO/ho chimeric genes.;
RL Curr. Genet. 28:367-373 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RA Rasmussen S.W.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP REVIEW.
RX MEDLINE=91171884; PubMed=2005783;
RA Herskowitz I., Jensen R.;
RT "Putting the HO gene to work: practical uses for mating-type
RT switching";
RL Meth. Enzymol. 194:132-146 (1991).

CC -1- FUNCTION: INITIATION OF MATING TYPE INTERCONVERSION. THIS PROTEIN
CC IS A SITE-SPECIFIC ENDONUCLEASE THAT CLEAVES A SITE IN THE MAT
CC LOCUS ON CHROMOSOME III. THE DOUBLE-STRAND BREAK IS FOLLOWED BY A
CC UNIDIRECTIONAL GENE CONVERSION EVENT THAT REPLACES THE INFORMATION
CC AT THE MAT LOCUS BY INFORMATION COPIED FROM EITHER OF THE TWO
CC HOMOLOGOUS LOCI (HMR AND HML) THAT RESIDE AT THE EXTREMITY OF THE
CC CHROMOSOME III. ENDONUCLEASE EXPRESSION TAKES PLACE IN LATE G1
CC JUST BEFORE CELLS ENTER S PHASE.
CC -1- MISCELLANEOUS: THE METAL-BINDING DOMAIN FORM ZINC-FINGERS THAT ARE
CC INVOLVED IN BINDING OF THE DNA.
CC -1- SIMILARITY: TO YEAST VMA1-DETERMINED ENDONUCLEASE (VDE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M14678; AAA34683.1; -;
CC EMBL; X90957; CAA62447.1; -;
CC EMBL; Z74275; CAA98806.1; -;
CC PIR; A25390; A25390.
CC HSSP; P17255; 1VDE.
CC SGD; S0002386; HO.
CC InterPro; IPR003587; HintN.
CC SMART; SM00306; HintN; 1.
CC Hydrolase; Endonuclease; Zinc; Zinc-finger; DNA-binding.
CC SIMILAR 325 334

FT CONFLICT 189 189 A -> T (IN REF. 1).
FT CONFLICT 223 223 S -> G (IN REF. 1).
FT CONFLICT 405 405 S -> L (IN REF. 1).
FT CONFLICT 475 475 L -> H (IN REF. 1).
SQ SEQUENCE 586 AA; 66089 MW; 95771394D177823A CRC64;

Query Match 42.9%; Score 6; DB 1; Length 586;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLDSD 7
Db 328 GLDSD 333

RESULT 8
ID SYFB_MYCPN STANDARD; PRT; 805 AA.
AC P75563;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.20) (PHENYLALANINE--
DE TRNA LIGASE BETA CHAIN) (PHERS).
GN PHET OR MPN106 OR MP048.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449 (1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) = AMP +
CC PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC CHAIN FAMILY. SUBFAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000006; AAB95696.1; -.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 805 AA; 91713 MW; 937B7614E789547A CRC64;

Query Match 42.9%; Score 6; DB 1; Length 805;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDS 6
|||||
Db 560 VGLIDS 565

RESULT 9
AMPN_LACLA STANDARD; PRT; 848 AA.
AC Q48636;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE AMINOPEPTIDASE N (EC 3.4.11.2) (LYSL AMINOPEPTIDASE) (LYS-AP)
DE (ALANINE AMINOPEPTIDASE).
GN DEIN
PEPN.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/clostridium group; Streptococcaceae;
OC Lactococcus
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=YRC001;
RA Tsukasaki F., Motoshima H., Minagawa E., Kaminogawa S.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: AMINOPEPTIDASE WITH BROAD SUBSTRATE SPECIFICITY TO
CC SEVERAL PEPTIDES. IT HAS MORE AFFINITY FOR OLIGOPEPTIDES THAN FOR
CC DIPEPTIDES. IT PLAYS AN ESSENTIAL ROLE IN THE METABOLISM, IT MAY
CC BE INVOLVED IN NITROGEN SUPPLY OR PROTEIN TURNOVER.
CC -1- COFACTOR: BINDS ONE ZINC ION.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. IT MAY BE SECRETED THROUGH
CC AN UNKNOWN MECHANISM.
CC -1- SIMILARITY: BELONGS TO THE PEPN SUBFAMILY.
CC ALSO KNOWN AS THE PEPN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D38040; BAA07234.1; -.
DR MEROPS; M01.002; -.
DR InterPro; IPR001930; Aladiptase.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc.
FT INIT_MET 0
FT METAL 293 293 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 294 294 BY SIMILARITY.
FT METAL 297 297 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 316 316 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 381 381 PROTON DONOR (POTENTIAL).
SQ SEQUENCE 848 AA; 96434 MW; 6EDEL716CF2B1185 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 848;
Best Local Similarity 100.0%; Pred. No. 24;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 DYQGQL 12
|||||
Db 526 DYQGQL 531

RESULT 10
VATA_YEAST STANDARD; PRT; 1071 AA.
AC P17255; O74301;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A (EC 3.6.1.34) (V-ATPASE A
DE SUBUNIT) [CONTAINS: ENDONUCLEASE PI-SCEI (EC 3.1.-.-) (VMAL-DERIVED
DE ENDONUCLEASE) (VDE) (SEE VMA INTEIN)].
GN (VMAL OR TFP1 OR CLS8 OR YDL185W OR D1286) AND VDE.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=X2180-1A;
RX MEDLINE=90216698; PubMed=2139027;
RA Hirata R., Ohsumi Y., Nakano A., Kawasaki H., Suzuki K., Anraku Y.;
RT "Molecular structure of a gene, VMAL, encoding the catalytic subunit
RT of H(+)-translocating adenosine triphosphatase from vacuolar
RT membranes of Saccharomyces cerevisiae.";
RL J. Biol. Chem. 265:6726-6733(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96021607; PubMed=8533471;
RA Verhasselt P., Voet M., Volckaert G.;
RT "New open reading frames, one of which is similar to the nlfv gene of
RT Azotobacter vinelandii, found on a 12.5 kbp fragment of chromosome IV
RT of Saccharomyces cerevisiae.";
RL Yeast 11:961-969(1995).
RN [3]
RP SEQUENCE OF 41-1071 FROM N.A.
RX MEDLINE=89096895; PubMed=2905423;
RA Shih C.K., Wagner R., Feinstein S., Kanik-Ennulat C., Neff N.;
RT "A dominant trifluoperazine resistance gene from Saccharomyces
RT cerevisiae has homology with F0F1 ATP synthase and confers calcium-
RT sensitive growth.";
RL Mol. Cell. Biol. 8:3094-3103(1988).
RN [4]
RP SEQUENCE OF 1-34 FROM N.A.
RC STRAIN=W303-1A;
RA Ronne H.E.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
RN [5]
RP PROTEIN SPLICING.
RX MEDLINE=91047969; PubMed=2146742;
RA Kane P.M., Yamashiro C.T., Wolczyk D.F., Neff N., Goebel M.,
RA Stevens T.H.;
RT "Protein splicing converts the yeast TFP1 gene product to the 69-kD
RT subunit of the vacuolar H(+)-adenosine triphosphatase.";
RL Science 250:651-657(1990).
RN [6]
RP MUTAGENESIS OF CYS-284 AND CYS-738.
RX MEDLINE=93038691; PubMed=1417861;
RA Hirata R., Anraku Y.;
RT "Mutations at the putative junction sites of the yeast VMAL protein,
RT the catalytic subunit of the vacuolar membrane H(+)-ATPase, inhibit
RT its processing by protein splicing.";
RL Biochem. Biophys. Res. Commun. 188:40-47(1992).
RN [7]
RP SELF-SPLICING MECHANISM.
RX MEDLINE=93285129; PubMed=8508780;
RA Cooper A.A., Chen Y.-J., Lindorfer M.A., Stevens T.H.;

"protein splicing of the yeast TFPI intervening protein sequence: a model for self-excision."
EMBL J. 12:2575-2583(1993).
[8]
FUNCTION OF VDE.
MEDLINE-92269953; PubMed-1534148;
Gimble F.S., Thorner J.;
"Homing of a DNA endonuclease gene by meiotic gene conversion in Saccharomyces cerevisiae."
Nature 357:301-306(1992).
[9]
REVIEW.
Grivell L.A.;
"Homing in on an endosymbiotic endonuclease."
Curr. Biol. 2:450-452(1992).
[10]
X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF PI-SCE I.
MEDLINE-97304384; PubMed-9160747;
Duan X., Gimble F.S., Quiococho F.A.;
"Crystal structure of Pi-SceI, a homing endonuclease with protein splicing activity."
Cell 89:555-564(1997).
[11]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF PI-SCE I.
MEDLINE-20112838; PubMed-10644733;
Hu D., Crist M., Duan X., Quiococho F.A., Gimble F.S.;
"Probing the structure of the Pi-SceI-DNA complex by affinity cleavage and affinity photocross-linking."
J. Biol. Chem. 275:2705-2712(2000).
-1- FUNCTION: CATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF VACUOLAR ATPASE. V-ATPASE VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS. IT IS AN ELECTROGENIC PROTON PUMP THAT GENERATES A PROTON MOTIVE FORCE OF 180 MV, INSIDE POSITIVE AND ACIDIC. IN THE VACUOLAR MEMBRANE VESICLES. IT MAY PARTICIPATE IN MAINTENANCE OF CYTOPLASMIC CA(2+) HOMEOSTASIS. THIS IS A CATALYTIC SUBUNIT.
-1- FUNCTION: PI-SCEI IS AN ENDONUCLEASE THAT CAN CLEAVE AT A SITE PRESENT IN A VMA1 ALLELE THAT LACKS THE DERIVED ENDONUCLEASE SEGMENT OF THE OPEN READING FRAME. CLEAVAGE AT THIS SITE ONLY OCCURS DURING MEIOSIS AND INITIATES "HOMING" A GENETIC EVENT THAT CONVERTS VMA1 ALLELES LACKING VDE INTO ONE THAT CONTAINS IT.
-1- SUBUNIT V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B, C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).
-1- SUBCELLULAR LOCATION: MEMBRANES OF VARIOUS INTRACELLULAR ACIDIC COMPARTMENTS.
-1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES A POST-TRANSLATIONAL EXCISION OF THE VDE REGION FOLLOWED BY PEPTIDE LIGATION.
-1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
-1- SIMILARITY: VDE IS HIGHLY SIMILAR TO ENDONUCLEASE HO.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: J05409; AAA34664.1; -
EMBL: X83276; CAA58261.1; -
EMBL: 274233; CAA98760.1; -
EMBL: 274233; CAA98761.1; ALT_SEQ.
EMBL: M21609; AAB63978.1; -
EMBL: X58857; CAA41657.1; -
PIR: A35746; PXBYVA.
PDB: 1VDE; 08-APR-98.
PDB: 1DFA; 08-DEC-99.
REBASE: 2615; PI-SceI.
SGD: S0002344; TFPI.
InterPro: IPR000793; ATPase_AB_C.

DR InterPro: IPR000194; ATPase_alpha_beta.
DR InterPro: IPR003586; HintC.
DR InterPro: IPR003587; HintN.
DR InterPro: IPR002203; Intein.
DR Pfam: PF00006; ATP-synt_ab; 1.
DR Pfam: PF00306; ATP-synt_ab_C; 1.
DR SMART: SM00305; HintC; 1.
DR SMART: SM00306; HintN; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE: PS00881; PROTEIN_SPLICING; 1.
KW ATP synthesis; Hydrogen ion transport; Hydrolase; ATP-binding;
KW Nuclease; Endonuclease; DNA-binding; Autocatalytic cleavage;
KW Protein splicing; Intron homing; 3D-structure;
FT CHAIN 1 283 VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT
FT A, 1ST PART.
FT CHAIN 284 737 ENDONUCLEASE PI-SCEI.
FT CHAIN 738 1071 VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT
FT NP_BIND 257 264 ATP (BY SIMILARITY).
FT SIMILAR 601 610 CONSERVED MOTIF FOR MITOCHONDRIAL MRNA
FT MUTAGEN 284 284 C->S: INHIBITS SPLICING.
FT MUTAGEN 738 738 C->S: INHIBITS SPLICING.
FT CONFLICT 875 875 G -> D (IN REF. 3).
SQ SEQUENCE 1071 AA; 118636 MW; 2A4C65D2F59426FD CRC64;

Query Match 42.9%; Score 6; DB 1; Length 1071;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLDSD 7
Db 604 GLDSD 609
|111111

RESULT 11
PAP2_SPVKA STANDARD; PRT; 69 AA.
ID PAP2_SPVKA
AC P23334; 1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POLY(A) POLYMERASE REGULATORY SUBUNIT (PAP) (VP39) (FRAGMENT).
GN SWF9.
OS Swinepox virus (strain Kasza) (SPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Suipoxvirus.
OX NCBI_TaxID=10277;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91196265; PubMed-1840707;
RA Schnitzlein W.M., Tripathy D.N.;
RT "Identification and nucleotide sequence of the thymidine kinase gene of swinepox virus."
RL Virology 181:727-732(1991).
CC -1- FUNCTION: POLYMERASE THAT CREATES THE 3' POLY(A) TAIL OF MRNA'S.
CC THE REGULATORY SUBUNIT BINDS TO POLY(A) BUT HAS NO CATALYTIC ACTIVITY.
CC -1- SUBUNIT: HETERODIMER OF VP55 (CATALYTIC) AND VP39 (REGULATORY).
CC -1- SIMILARITY: TO OTHER POXVIRUSES VP39.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M59931; AAA47893.1; -
PIR: D37949; D37949.
HSSP: P07617; LAV6.

```
DR InterPro: IPR000176; PolA_polym_reg.
DR Pfam: PF01358; PARP_regulatory; 1.
KW Transferase; Transcription.
FT NON_TER 69
SQ SEQUENCE 69 AA; 8069 MW; C6830F285030213E CRC64;

Query Match 35.7%; Score 5; DB 1; Length 69;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YQGL 12
DB 36 YQGL 40

RESULT 12
YGAV_ECOLI STANDARD; PRT; 99 AA.
AC P77295;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR YGAV.
GN YGAV OR B2657.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitaawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Washimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
CC -1- SIMILARITY: BELONGS TO THE ARSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000351; AAC75714.1; -
CC EMBL; D90890; BAA16530.1; -
CC EMBL; D90891; BAA16534.1; -
CC EcoGene; EG13524; ygav.
CC InterPro; IPR001845; HTH_ARSR.
CC Pfam; PF01022; HTH_5; 1.
CC SMART; SMO0418; HTH_ARSR; 1.
CC PROSITE; PS00846; HTH_ARSR_FAMILY; FALSE_NEG.
CC Hypothetical protein; Transcription regulation; DNA-binding;

KW Complete proteome.
FT DNA_BIND 41 60 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 99 AA; 10596 MW; 93B9B32C62CC0223 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDS 6
DB 66 GLIDS 70

RESULT 13
YJZ2_YEAST STANDARD; PRT; 128 AA.
AC P47093;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 14.5 KDA PROTEIN IN MER2-BNA1 INTERGENIC REGION.
GN YJR022W OR J1464 OR YUR83.16.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA de Haan M., Smits P.H.M., Grivell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 66-128 FROM N.A.
RX MEDLINE=96109930; PubMed=8619316;
RA Zagulski M., Babinska B., Gromadka R., Migdalski A., Rytka J.,
RA Sulicka J., Herbert C.J.;
RT "The sequence of 24.3 kb from chromosome X reveals five complete open
RT reading frames, all of which correspond to new genes, and a tandem
RT insertion of a Tyl transposon.";
RL Yeast 11:1179-1186(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X87611; CAA60945.1; -
CC EMBL; Z49522; CAA89547.1; -
CC SGD; S0003783; YJR022W.
CC InterPro; IPR001163; snRNP_Sm.
CC Pfam; PF01423; Sm; 1.
CC KW Hypothetical protein.
SQ SEQUENCE 128 AA; 14518 MW; F05592F393BEF3DF CRC64;

Query Match 35.7%; Score 5; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLID 5
DB 83 VGLID 87

RESULT 14
DUT_CHVP1 STANDARD; PRT; 141 AA.
ID DUT_CHVP1
AC O41033;
DT 15-DEC-1998 (Rel. 37, Created)
```

DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DROXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPASE) (DUTP PYROPHOSPHATASE).
GN A551L.
OS Paramacium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9802962; PubMed=9356347;
RA Li Y., Lu Z., Sun L., Ropp S., Kutish G.F., Rock D.L., van Etten J.L.;
RT "Analysis of 74 kb of DNA located at the right end of the 330-kb
RL chlorella virus PBCV-1 genome.";
RL Virology 237:360-377(1997).
CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
CC URACIL CANNOT BE INCORPORATED INTO DNA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U42580; AAC96912.1; -
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
KW Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 141 AA; 14879 MW; 3D0734C9C9BD70B8 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 IDS DY 8
Db 79 IDS DY 83

RESULT 15
DUT_SPVKA
ID DUT_SPVKA STANDARD; PRT; 142 AA.
AC P32208;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DROXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPASE) (DUTP PYROPHOSPHATASE).
GN Clil.
OS Swinepox virus (strain Kasza) (SPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Suipoxvirus.
OX NCBI_TaxID=10277;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94069924; PubMed=8249275;
RA Massung R.F., Jayarama V., Moyer R.W.;
RA "DNA sequence analysis of conserved and unique regions of swinepox
RT virus: identification of genetic elements supporting phenotypic
RT observations including a novel G protein-coupled receptor
RT homologue.";
RL Virology 197:511-528(1993).
CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
CC PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
CC AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT

CC URACIL CANNOT BE INCORPORATED INTO DNA.
CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PROTEASE-LIKE PROTEIN
CC (PSEUDOPROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L22013; AAC37860.1; -
DR HSP: P06968; 1DUD.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
KW Hydrolase; Nucleotide metabolism.
SQ SEQUENCE 142 AA; 15750 MW; 13BCDF9C64AD9635 CRC64;
Query Match 35.7%; Score 5; DB 1; Length 142;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 IDS DY 8
Db 80 IDS DY 84
Search completed: January 31, 2002, 13:39:23
Job time: 83 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:21 ; Search time 46.78 seconds

(without alignments)

10.973 Million cell updates/sec

Title: US-08-957-709-77

Perfect score: 14

Sequence: 1 AHRIDPGWGCIVL 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	193	1 DCD_ECOLI	P28248 escherichia
2	8	57.1	193	1 DCD_BUCAL	P57209 buchnera ap
3	8	57.1	194	1 DCD_PASMU	P57891 pasteurella
4	8	57.1	195	1 DCD_HAEN	P44534 haemophilus
5	7	50.0	193	1 DCD_BUCAP	O9zhd8 buchnera ap
6	6	42.9	141	1 Y085_BORBU	O51112 borrelia bu
7	6	42.9	340	1 CEUF_VZVD	P09261 varicella-z
8	6	42.9	733	1 YFJZ_YEAST	P43602 saccharomyc
9	6	42.9	2444	1 NTC1_HUMAN	P46531 homo sapien
10	5	35.7	78	1 IF1C_MARPO	P12134 marchantia
11	5	35.7	107	1 FER_PSALA	P34806 psalteriomo
12	5	35.7	132	1 XJTW_ECOLI	P39394 escherichia
13	5	35.7	141	1 HB44_XENLA	P06637 xenopus lae
14	5	35.7	142	1 HBAC_ANGAN	P80726 anguilla an
15	5	35.7	191	1 PCD6_HUMAN	O75340 homo sapien
16	5	35.7	191	1 PCD6_MOUSE	P12815 mus musculu
17	5	35.7	216	1 CSGD_SALTY	O54294 salmonella
18	5	35.7	227	1 N4PC_RHOSH	O53178 rhodobacter
19	5	35.7	245	1 Y4PL_RHISN	P55617 rhizobium s
20	5	35.7	249	1 COBN_RHOSH	O53138 rhodococcus
21	5	35.7	268	1 Y084_MYCTU	O53191 mycobacteri
22	5	35.7	294	1 YAFJ_HAEN	P44098 haemophilus
23	5	35.7	298	1 EFTS_MYCPN	P78009 mycoplasma
24	5	35.7	323	1 Y370_MYCGE	P47610 mycoplasma
25	5	35.7	326	1 Y370_MYCPN	P75230 mycoplasma
26	5	35.7	362	1 ML1B_HUMAN	P49286 homo sapien
27	5	35.7	410	1 B3_USTMA	P22017 ustilago ma
28	5	35.7	470	1 RHSA_RHIME	O9z3r2 rhizobium m
29	5	35.7	512	1 ER24_SEPLY	O13597 septoria ly
30	5	35.7	549	1 TEGU_HCMV	O7387 human cytom
31	5	35.7	663	1 RGPI_YEAST	P16864 saccharomyc
32	5	35.7	668	1 COAT_FCVCF6	P27404 feline cali
33	5	35.7	668	1 COAT_FCVF4	P27405 feline cali

ALIGNMENTS

RESULT 1

ID	DCD_ECOLI	STANDARD;	PRT;	193 AA.
AC	P28248;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP DEAMINASE).			
GN	DCD OR DUS OR PAXA OR B2065.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.			
RC	STRAIN=K12;			
RX	MEDLINE=92380941; PubMed=1324907;			
RA	Wang L., Weiss B.;			
RT	"dcd (dCTP deaminase) gene of Escherichia coli: mapping, cloning, sequencing, and identification as a locus of suppressors of lethal dut (dUTPase) mutations."			
RL	J. Bacteriol. 174:5647-5653(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Science 277:1453-1474(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97251358; PubMed=9097040;			
RA	Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima H., Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;			
RT	"A 460-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 40.1-50.0 min region on the linkage map."			
RL	DNA Res. 3:379-392(1996).			
CC	-1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).			
CC	-1- SUBUNIT: HOMOTETRAMER (PROBABLE).			
CC	-1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/).			

P27406 feline cali
O9zky0 helicobacte
P50610 helicobacte
O46861 escherichia
Q18600 caenorhabdl
P52550 gallus gall
P15558 pseudomonas
P39920 coxiella bu
O9rlv6 mus musculu
P25733 escherichia
Q11042 mycobacteri
O9p0k1 homo sapien

```

or send an email to license@lsb-sib.ch).
-----
CC DR EMBL; M90069; AAA23669.1; -
CC DR EMBL; AE000296; AAC75126.1; -
CC DR EMBL; D90844; BAA15918.1; -
CC DR EMBL; D90845; BAA15923.1; -
CC DR PIR; A42940; A42940.
CC DR EcoGene; EG11418; dcd.
CC DR InterPro; IPR003232; dCTP_deaminase.
CC DR InterPro; IPR001428; dUTPase.
CC DR Pfam; PF00692; dUTPase; 1.
CC DR ProDom; PD004900; dCTP_deaminase; 1.
CC DR Hydrolase; Complete proteome.
CC KW SEQUENCE 193 AA; 21249 MW; B0044051ADE7F919 CRC64;
SQ
-----
Query Match 100.0%; Score 14; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AHRIDPCWGCIVL 14
DB 124 AHRIDPCWGCIVL 137
-----
RESULT 2
ID DCD_BUCAI STANDARD; PRT; 193 AA.
AC P57209;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR Bui108.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@lsb-sib.ch).
-----
CC EMBL; AP001118; BAB12827.1; ALT INIT.
CC InterPro; IPR003232; dCTP_deaminase.
CC InterPro; IPR001428; dUTPase.
CC Pfam; PF00692; dUTPase; 1.
CC ProDom; PD004900; dCTP_deaminase; 1.
CC Hydrolase; Complete proteome.
CC KW SEQUENCE 193 AA; 21951 MW; DE7DD7FD961FC9FA CRC64;
SQ
-----
Query Match 57.1%; Score 8; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AHRIDPCW 8
DB 124 AHRIDPCW 8
-----
or send an email to license@lsb-sib.ch).
-----
CC DR EMBL; M90069; AAA23669.1; -
CC DR EMBL; AE000296; AAC75126.1; -
CC DR EMBL; D90844; BAA15918.1; -
CC DR EMBL; D90845; BAA15923.1; -
CC DR PIR; A42940; A42940.
CC DR EcoGene; EG11418; dcd.
CC DR InterPro; IPR003232; dCTP_deaminase.
CC DR InterPro; IPR001428; dUTPase.
CC DR Pfam; PF00692; dUTPase; 1.
CC DR ProDom; PD004900; dCTP_deaminase; 1.
CC DR Hydrolase; Complete proteome.
CC KW SEQUENCE 193 AA; 21249 MW; B0044051ADE7F919 CRC64;
SQ
-----
Query Match 100.0%; Score 14; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.4e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AHRIDPCWGCIVL 14
DB 124 AHRIDPCWGCIVL 137
-----
RESULT 2
ID DCD_BUCAI STANDARD; PRT; 193 AA.
AC P57209;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR Bui108.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@lsb-sib.ch).
-----
CC EMBL; AP001118; BAB12827.1; ALT INIT.
CC InterPro; IPR003232; dCTP_deaminase.
CC InterPro; IPR001428; dUTPase.
CC Pfam; PF00692; dUTPase; 1.
CC ProDom; PD004900; dCTP_deaminase; 1.
CC Hydrolase; Complete proteome.
CC KW SEQUENCE 193 AA; 21951 MW; DE7DD7FD961FC9FA CRC64;
SQ
-----
Query Match 57.1%; Score 8; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AHRIDPCW 8
DB 124 AHRIDPCW 8
-----

```

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
CC -I- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32699; AAC21805.1; -
CC TIGR: H10133; -
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dUTPase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 195 AA; 21616 MW; 39E5C2370A6C1CB CRC64;

Query Match 57.1%; Score 8; DB 1; Length 195;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHRIDPGW 8
Db 124 AHRIDPGW 131
|||||

RESULT 5
DCD_BUCAP STANDARD; PRT; 193 AA.
ID DCD_BUCAP
AC Q9ZHD8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DSOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98440331; Pubmed=9767718;
RA Clark M.A., Baumann L., Baumann P.;
RT "Buchnera aphidicola (Aphid endosymbiont) contains genes encoding
RT enzymes of histidine biosynthesis.";
RL Curr. Microbiol. 37:356-358(1998).
CC -I- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -I- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF067228; AAC97363.1; -
DR InterPro: IPR003232; dCTP_deaminase.

DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hydrolase.
SQ SEQUENCE 193 AA; 21871 MW; 8E19D4580C7C55E8 CRC64;

Query Match 50.0%; Score 7; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 HRIDPGW 8
Db 125 HRIDPGW 131
|||||

RESULT 6
Y085_BORBU STANDARD; PRT; 141 AA.
ID Y085_BORBU
AC O51112;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN BB0085.
GN BB0085.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; Pubmed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Wattney L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001121; AAC66475.1; -
DR TIGR; BB0085; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 141 AA; 16214 MW; 1F42B5CABFAA690F CRC64;

Query Match 42.9%; Score 6; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGCIVL 14
Db 58 SGCIVL 63
|||||

RESULT 7
CELLF_VZVD STANDARD; PRT; 340 AA.
ID CELLF_VZVD
AC P09261;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CELL FUSION PROTEIN PRECURSOR.
 GN 5.
 OS Varicella-zoster virus (strain Dumas) (VZV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=10338;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86306657; PubMed=3018124;
 RA Davison A.J., Scott J.E.;
 RT "The complete DNA sequence of varicella-zoster virus."
 RL J. Gen. Virol. 67:1759-1816(1986).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X04370; CAA27888.1;
 DR EMBL: E27212; MMBE5
 DR InterPro: IPR002567; Fusion_gly_K.
 DR Pfam: PF01621; Fusion_gly_K; 1.
 KW Fusion protein; Transmembrane; Signal.
 FT SIGNAL ?
 FT CHAIN ? 340 CELL FUSION PROTEIN.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 340 AA; 38576 MW; 0387FE00EC39C946 CRC64;
 SQ
 Query Match 42.9%; Score 6; DB 1; Length 340;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 SGCIVL 14
 DB 226 SGCIVL 231
 RESULT 8
 YFJ2_YEAST
 ID YFJ2_YEAST STANDARD; PRT; 733 AA.
 AC P43602;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 79.7 KDA PROTEIN IN FABL-PES4 INTERGENIC REGION.
 GN YFRO22W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=S288C / AB972;
 RC MEDLINE=95400232; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagihara H., Shibata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Tamazaki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 RT Saccharomyces cerevisiae."
 RL Nat. Genet. 10:261-268(1995).
 CC -!- SIMILARITY: TO YEAST ROD1, YGR068C AND TO S.POMBE SPAC31A2.12 AND
 CC SPAC8A4.13C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D50617; BAA09261.1;
 DR EMBL: S0001918; YFRO22W.
 KW Hypothetical protein.
 FT DOMAIN 80 87 POLY-SER.
 FT DOMAIN 124 128 POLY-SER.
 FT DOMAIN 423 427 POLY-GLU.
 FT DOMAIN 679 684 POLY-SER.
 SQ SEQUENCE 733 AA; 79708 MW; 2BC6F724B9806A5 CRC64;
 Query Match 42.9%; Score 6; DB 1; Length 733;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 SGCIVL 14
 DB 43 SGCIVL 48
 RESULT 9
 NTC1_HUMAN
 ID NTC1_HUMAN STANDARD; PRT; 2444 AA.
 AC P46531;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLATION -
 DE ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).
 GN NOTCH1 OR TAN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91347367; PubMed=1831692;
 RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
 RA Smith S.D., Sklar J.;
 RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
 RT chromosomal translocations in T lymphoblastic neoplasms."
 RL Cell 66:649-661(1991).
 CC -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN
 CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
 CC IN SOME T-CELL NEOPLASMS.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
 CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
 CC IS FOUND MAINLY IN LYMPHOID TISSUES.
 CC -!- SIMILARITY: HIGH WITH OTHER NOTCH-TYPE PROTEINS.
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M73980; AAA60614.1;
 DR EMBL: P00740; 11XA.
 DR MIM; 190198;
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.

```

DR InterPro: IPR000800; Notch.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 36.
DR Pfam: PF00066; notch; 3.
DR SMART: SM00248; ank; 5.
DR SMART: SM00179; EGF_CA; 22.
DR SMART: SM00001; EGF_Like; 15.
DR SMART: SM00004; NL; 2.
DR PROSITE: PS00088; ANK_REPEAT; 4.
DR PROSITE: PS0297; ANK_REPEAT; 1.
DR PROSITE: PS00110; ASX_HYDROXYL; 20.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS0186; EGF_2; 26.
DR PROSITE: PS01187; EGF_CA; 18.
DR PROSITE: PS01187; EGF_CA; 18.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 >2444 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.
FT DOMAIN 19 1736 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1737 1757 POTENTIAL.
FT DOMAIN 1758 >2444 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 58 EGF-LIKE 1.
FT DOMAIN 59 99 EGF-LIKE 2.
FT DOMAIN 102 139 EGF-LIKE 3.
FT DOMAIN 140 176 EGF-LIKE 4.
FT DOMAIN 178 216 EGF-LIKE 5.
FT DOMAIN 218 255 EGF-LIKE 6.
FT DOMAIN 257 293 EGF-LIKE 7.
FT DOMAIN 295 333 EGF-LIKE 8.
FT DOMAIN 335 371 EGF-LIKE 9.
FT DOMAIN 372 410 EGF-LIKE 10.
FT DOMAIN 412 450 EGF-LIKE 11.
FT DOMAIN 452 488 EGF-LIKE 12.
FT DOMAIN 490 526 EGF-LIKE 13.
FT DOMAIN 528 564 EGF-LIKE 14.
FT DOMAIN 566 601 EGF-LIKE 15.
FT DOMAIN 603 639 EGF-LIKE 16.
FT DOMAIN 641 676 EGF-LIKE 17.
FT DOMAIN 678 714 EGF-LIKE 18.
FT DOMAIN 716 751 EGF-LIKE 19.
FT DOMAIN 753 789 EGF-LIKE 20.
FT DOMAIN 791 827 EGF-LIKE 21.
FT DOMAIN 829 868 EGF-LIKE 22.
FT DOMAIN 870 906 EGF-LIKE 23.
FT DOMAIN 908 944 EGF-LIKE 24.
FT DOMAIN 946 982 EGF-LIKE 25.
FT DOMAIN 984 1020 EGF-LIKE 26.
FT DOMAIN 1022 1058 EGF-LIKE 27.
FT DOMAIN 1060 1096 EGF-LIKE 28.
FT DOMAIN 1098 1144 EGF-LIKE 29.
FT DOMAIN 1146 1182 EGF-LIKE 30.
FT DOMAIN 1184 1220 EGF-LIKE 31.
FT DOMAIN 1222 1266 EGF-LIKE 32.
FT DOMAIN 1268 1306 EGF-LIKE 33.
FT DOMAIN 1308 1347 EGF-LIKE 34.
FT DOMAIN 1349 1385 EGF-LIKE 35.
FT DOMAIN 1388 1427 EGF-LIKE 36.
FT REPEAT 1446 1481 LIN/NOTCH 1.
FT REPEAT 1482 1523 LIN/NOTCH 2.
FT REPEAT 1524 1563 LIN/NOTCH 3.
FT REPEAT 1528 1557 ANK 1.
FT REPEAT 1561 1591 ANK 2.
FT REPEAT 1595 2024 ANK 3.
FT REPEAT 2028 2057 ANK 4.
FT REPEAT 2061 2090 ANK 5.
FT DOMAIN 1576 1579 POLY-VAL.
FT DOMAIN 1662 1665 POLY-ARG.
FT DOMAIN 1729 1732 POLY-PRO.
FT DOMAIN 1741 1744 POLY-ALA.
FT DOMAIN 1902 1905 POLY-GLU.
FT DOMAIN 2260 2263 POLY-GLY.
FT DOMAIN 2404 2407 POLY-GLN.
FT DOMAIN 2411 2418 POLY-PRO.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 233 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 660 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 855 BY SIMILARITY.
FT DISULFID 857 867 BY SIMILARITY.
FT DISULFID 874 885 BY SIMILARITY.
FT DISULFID 879 894 BY SIMILARITY.
FT DISULFID 896 905 BY SIMILARITY.
FT DISULFID 912 923 BY SIMILARITY.
FT DISULFID 934 943 BY SIMILARITY.
FT DISULFID 948 959 BY SIMILARITY.

```

```

FT FT DISULFID 993 1008 BY SIMILARITY.
FT FT DISULFID 1010 1019 BY SIMILARITY.
FT FT DISULFID 1026 1037 BY SIMILARITY.
FT FT DISULFID 1031 1046 BY SIMILARITY.
FT FT DISULFID 1048 1057 BY SIMILARITY.
FT FT DISULFID 1064 1075 BY SIMILARITY.

Query Match 42.9%; Score 6; DB 1; Length 2444;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGWSG 10
DB 742 DPGWSG 747

RESULT 10
ID IFIC_MARPO STANDARD; PRT; 78 AA.
AC P12134; 1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSLATION INITIATION FACTOR IF-1, CHLOROPLAST.
GN INF.
OS Marchantia polymorpha (Liverwort).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Marchantiophyta;
OC Marchantiales; Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89068687; PubMed=3199436;
RA Fukuzawa H., Kohchi T., Sano T., Shirai H.,
Ozeki H., Ohyama K.;
RT "Structure and organization of Marchantia polymorpha chloroplast
genome. III. Gene organization of the large single copy region from
rbcl to trn(CAU).";
RL J. Mol. Biol. 203:333-351(1988).
RN [2]
RN COMPLETE GENOME.
RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
Umesono K., Shiki Y., Takeuchi M., Chang Z., Ozeki H.,
Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574(1986).
CC -1- FUNCTION: NO SPECIFIC FUNCTION HAS SO FAR BEEN ATTRIBUTED TO THIS
INITIATION FACTOR; HOWEVER, IT SEEMS TO STIMULATE MORE OR LESS ALL
THE ACTIVITIES OF THE OTHER TWO INITIATION FACTORS, IF-2 AND IF-3.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- SIMILARITY: BELONGS TO THE IF-1 FAMILY.
CC -1- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X04465; CAA28120.1;
DR PIR; A05008; A05008.
DR HSP; P02998; 1AH9.
DR Mendel; 2310; MARPO; info:1.
DR InterPro; IPR003029; S1.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00316; S1; 1.
KW Initiation factor; Protein biosynthesis; Chloroplast.
FT DOMAIN 22 71 SI MOTIF
FT SEQUENCE 78 AA; 8978 MW; B0C85D43EB5AFB9 CRC64;
OS Escherichia coli.

Query Match 35.7%; Score 5; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGCIV 13
DB 14 SGCIV 18

RESULT 12
ID YJW_ECOLI STANDARD; PRT; 132 AA.
AC P39394;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 14.6 KDA PROTEIN IN MCRB-HSDS INTERGENIC REGION (F132).
GN YJW OR B4347.
OS Escherichia coli.

Query Match 35.7%; Score 5; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGCIV 13
DB 14 SGCIV 18

RESULT 11
ID FER_PSALA STANDARD; PRT; 107 AA.
AC P34806; 1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE FERREDOXIN PRECURSOR.
OS Psalteriomonas lanterna.
OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae;
OC Psalteriomonas.
OX NCBI_TaxID=31290;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=NJMEGEN;
RX MEDLINE=94114557; PubMed=8286402;
RA Brul S., Veltman R.H., Lombardo M.C.P., Vogels G.D.;
RT "Molecular cloning of hydrogenosomal ferredoxin cDNA from the
anaerobic amoeboid flagellate Psalteriomonas lanterna.";
RL Biochim. Biophys. Acta 1183:544-546(1994).
CC -1- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER.
CC -1- SUBCELLULAR LOCATION: HYDROGENOSOMAL.
CC -1- SIMILARITY: TO 2FE-2S FERREDOXINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X74556; CAA52650.1;
DR Electron transport; Iron-sulfur; Hydrogenosome.
FT PROPEP 1 8
FT CHAIN 9 107 FERREDOXIN.
FT METAL 45 45 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 51 51 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 54 54 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 107 AA; 11179 MW; F033ACCB00134CC3 CRC64;

```

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=95334362; PubMed=7610040;

RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,

RA Blattner F.R.;

RT "Analysis of the Escherichia coli genome VI: DNA sequence of the

RT region from 92.8 through 100 minutes."

RL Nucleic Acids Res. 23:2105-2119(1995).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U14003; AAA97244.1; -.

DR EMBL: AE000505; AAC77303.1; -.

DR EcoGene; EG12584; yjiW.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 132 AA; 14576 MW; 3E53097CD17B0C62 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 132;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GCIVL 14

DB 84 GCIVL 88

RESULT 13

HBA4_XENLA

ID HBA4_XENLA STANDARD; PRT; 141 AA.

AC P06637;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-APR-1988 (Rel. 07, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE HEMOGLOBIN ALPHA-4 CHAIN (ALPHA-T4).

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85297748; PubMed=2993998;

RA Banville D., Williams J.G.;

RT "The pattern of expression of the Xenopus laevis tadpole alpha-globin

RT genes and the amino acid sequence of the three major tadpole alpha-

RT globin polypeptides."

RL Nucleic Acids Res. 13:5407-5421(1985).

CC -1- FUNCTION: THIS IS A LARVAL (TADPOLE) ALPHA-GLOBIN.

CC -1- POLYMORPHISM: ALPHA T4 MAY BE AN ALLELE OF ALPHA T3.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X02797; CAA26565.1; -.

DR PIR; B24338; B24338.

DR HSSP; P01922; IFDH.

DR InterPro; IPR002338; Alpha_haem.

DR InterPro; IPR000971; Globin.
 DR Pfam; PF00042; globin; 1.
 DR PRINTS; PR00612; ALPHAHAE.
 DR PROSITE; PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Respiratory protein; Erythrocyte.
 FT INIT_MET 0 0
 FT METAL 58 58 IRON (HEME DISTAL LIGAND).
 FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 141 AA; 15322 MW; 41E6580905BDF546 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 141;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RIDPG 7

DB 92 RIDPG 96

RESULT 14

HBAC_ANGAN

ID HBAC_ANGAN STANDARD; PRT; 142 AA.

AC P80726;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HEMOGLOBIN CATHODIC, ALPHA CHAIN.

OS Anguilla anguilla (European freshwater eel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguilloidei;

OC Anguillidae; Anguilla.

OX NCBI_TaxID=7936;

RN [1]

RP SEQUENCE.

RC TISSUE=Erythrocyte;

RX MEDLINE=95370199; PubMed=7642546;

RA Fago A., Carratore V., di Prisco G., Feuerlein R.J.,

RA Sottrup-Jensen L., Weber R.E.;

RT "The cathodic hemoglobin of Anguilla anguilla. Amino acid sequence

RT and oxygen equilibria of a reverse Bohr effect hemoglobin with high

RT oxygen affinity and high phosphate sensitivity."

RL J. Biol. Chem. 270:18997-18902(1995).

CC -1- FUNCTION: INVOLVED IN OXYGEN TRANSPORT FROM THE LUNG TO THE

CC VARIOUS PERIPHERAL TISSUES.

CC -1- SUBUNIT: Tetramer of two alpha chains and two beta chains.

CC -1- TISSUE SPECIFICITY: RED BLOOD CELLS.

CC -1- MISCELLANEOUS: THIS FISH HAS TWO HEMOGLOBINS: CATHODIC AND ANODIC.

CC THE CATHODIC HB AND ANODIC HB DISPLAY SMALL AND LARGE BOHR EFFECTS

CC RESPECTIVELY. IN ADDITION, THE CATHODIC HB DISPLAYS A REVERSE BOHR

CC EFFECT AND APPRECIABLE PHOSPHATE EFFECTS.

CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.

CC HSSP; P02019; IOUW.

DR InterPro; IPR002338; Alpha_haem.

DR InterPro; IPR000971; Globin.

DR Pfam; PF00042; globin; 1.

DR PRINTS; PR00612; ALPHAHAE.

DR PROSITE; PS01033; GLOBIN; 1.

KW Heme; Oxygen transport; Respiratory protein; Erythrocyte; Acetylation.

FT MOD_RES 1 1 ACETYLATION (PROBABLE).

FT METAL 59 59 IRON (HEME DISTAL LIGAND).

FT METAL 88 88 IRON (HEME PROXIMAL LIGAND).

SQ SEQUENCE 142 AA; 15189 MW; F61A7B96A07A41CD CRC64;

Query Match 35.7%; Score 5; DB 1; Length 142;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RIDPG 7

DB 93 RIDPG 97

```
RESULT 15
PCD6_HUMAN STANDARD; PRT; 191 AA.
ID PCD6_HUMAN
AC 075340;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROGRAMED CELL DEATH PROTEIN 6 (PROBABLE CALCIUM-BINDING PROTEIN
DE ALG-2).
GN PCD6 OR ALG2.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ganjel J.K., D'Adamo L.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Urcelay E., Ibarreta D., Parrilla R., Ayuso M.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CALCIUM-BINDING PROTEIN REQUIRED FOR T CELL RECEPTOR-,
CC FAS-, AND GLUCOCORTICOID-INDUCED CELL DEATH. MAY MEDIATE CA(2+)-
CC REGULATED SIGNALS ALONG THE DEATH PATHWAY (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 5 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF035606; AAC27697.1;
DR EMBL: U58773; AAF14336.1;
DR MIM: 601057;
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; efhand; 5.
DR SMART: SM00054; EPH; 3.
DR PROSITE: PS00018; EF_HAND; 2.
KW Calcium-binding; Rept; Apoptosis.
FT CA_BIND 35 47 EF_HAND 1 (POTENTIAL).
FT CA_BIND 73 84 EF_HAND 2 (POTENTIAL).
FT CA_BIND 103 114 EF_HAND 3 (POTENTIAL).
FT DOMAIN 139 150 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT DOMAIN 169 180 ANCESTRAL CALCIUM SITE 5 (POTENTIAL).
SQ SEQUENCE 191 AA; 21868 MW; DOBS944CF3C696AD CRC64;

Query Match 35.7%; Score 5; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GCIVL 14
Db 154 GCIVL 158

Search completed: January 31, 2002, 13:39:22
Job time: 82 sec
```


Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	14	100.0	173	1	DCD_ACIAM	Q02103 acidilanus a
2	11	78.6	181	1	DCD_AERPE	O9yfa8 aeropyrum p
3	8	57.1	180	1	DCD_AQUAE	O67539 aquifex aeo
4	7	50.0	328	1	VGH_BPS13	P07933 bacterioph
5	7	50.0	984	1	T3RE_SALTY	P40815 salmonella
6	6	42.9	258	1	ARYL_FELCA	O62696 felis silve
7	6	42.9	268	1	DUT_PRVKA	O90030 pseudorabie
8	6	42.9	287	1	ARYL_CHICK	P12275 gallus gall
9	6	42.9	290	1	ARY1_CHICK	P13913 gallus gall
10	6	42.9	290	1	ARY1_HUMAN	P18440 homo sapien
11	6	42.9	290	1	ARY1_RABIT	P18605 oryctolagus
12	6	42.9	290	1	ARY2_MESAU	P50293 mesocricetu
13	6	42.9	290	1	ARY2_MOUSE	P50295 mus musculu
14	6	42.9	290	1	ARY2_RABIT	P11246 oryctolagus
15	6	42.9	290	1	ARY2_RAT	P50298 rattus norv
16	6	42.9	326	1	ODPB_MYCCE	P47515 mycoplasma
17	6	42.9	327	1	ODPB_ACHUA	P35488 acholeplasm
18	6	42.9	327	1	ODPB_MYCPN	P75391 mycoplasma
19	6	42.9	328	1	VGH_BPHFX	P03646 bacterioph
20	6	42.9	357	1	YFQB_SCHPO	Q10170 schizosacch
21	6	42.9	440	1	YHRI_YEAST	P38820 saccharomyc
22	6	42.9	525	1	SVK_DEIRA	Q9rxel deinococcus
23	6	42.9	578	1	VAC8_YEAST	P39968 saccharomyc
24	6	42.9	587	1	BARI_YEAST	P12630 saccharomyc
25	6	42.9	997	1	YNM3_YEAST	P35920 saccharomyc
26	6	42.9	1001	1	AHM5_ARATH	Q9s7j8 arabidopsis
27	6	42.9	1102	1	TRAA_RHISN	P55418 rhizobium s
28	6	42.9	1114	1	RH18_YEAST	Q12749 saccharomyc
29	6	42.9	2029	1	LAR_DROME	P16621 drosophila
30	5	35.7	772	1	NIFT_AZOVI	P09427 azotobacter
31	5	35.7	122	1	RL7_DEIRA	Q9rst0 deinococcus
32	5	35.7	130	1	RL32_THEAC	Q9his4 thermoplasm
33	5	35.7	133	1	PA23_OXYXS	P00616 oxvuranus s

RESULT 1			
DCD_ACIAM			
ID	DCD_ACIAM	STANDARD;	PRT; 173 AA.
AC	Q02103;		
CT	01-APR-1993	(Rel. 25, Created)	
DT	01-APR-1993	(Rel. 25, Last sequence update)	
DT	20-AUG-2001	(Rel. 40, Last annotation update)	
DE	PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP		
DE	DEAMINASE).		

ALIGNMENTS

DCL
 OS
 AC
 OC
 OX
 NCBI_TaxID=2283;
 RN [1]

RC	STRAIN=Lei 10 / DSM 3772;
RX	MEDLINE=93065206; PubMed=1437556;

"Molecular characterisation of a DNA ligase gene of the extremely thermophilic archaeon *Desulfurolobus ambivalens* shows close phylogenetic relationship to eukaryotic ligases.";
Nucleic Acids Res. 20:5389-5396(1992).

[2]
RN MEDLINE=95206934; PubMed=7899076;
KP Ouzounis C., Kyriides N., Sander C.;
RN "Novel protein families in archæean genomes.";
RT Nucleic Acids Res. 23:565-570(1995).
CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O = DUTP + NH(3).
CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

```
CC      EMBL; X63438; CAA45033.1; -.
DR      PIR; S26382; S26382.
DR      InterPro; IPR003232; dCTP_deaminase.
DR      InterPro; IPR001428; dUTPase.
DR      Pfam; PF00692; dUTPase; 1.
DR      ProDom; PD004900; dCTP_deaminase; 1.
DR      TrEMBL; P00000; dCTP_deaminase; 1.
```

KW Hydrolase.
SQ SEQUENCE 173 AA; 19858 MW; B4D922503CD4B25A CRC64;

```
Query Match      100.0%; Score 14; DB 1; Length 173;
Best Local Similarity 100.0%; Pred. No. 2.6e-08;
Matches 14; Conservative 0; Mismatches 0; Indels
```

1 p"IVDAGFEGOLTI 14

db 109 PTVDAGFEGOLTI 122

```

RESULT 2
DCD_AERPE          STANDARD;          PRT; 181 AA.
AC  Q3YFAB;
DT  20-AUG-2001 (Rel. 40, Created)
DT  20-AUG-2001 (Rel. 40, Last sequence update)
DE  20-AUG-2001 (Rel. 40, Last annotation update)
DE  PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE  DEAMINASE).
GN  DCD OR APE0333.
OS  Aeropyrum pernix.
OC  Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC  Aeropyrum.
OX  NCBI_TaxID=56636;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-K1;
RX  MEDLINE=59310339; PubMed=10382966;
RA  Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA  Jin-no K., Takahashi M., Sekine M., Baba S.-I., Nakai A., Kosugi H.,
RA  Hosovama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA  Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA  Yamazaki Y., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA  Nakamura Y., Nomura N., Sako I., Kikuchi H.; -thermophilic
RT  "Complete genome sequence of an aerobic hyper-thermophilic
RT  Crenarchaeon, Aeropyrum pernix K1.";
RL  DNA Res. 6:83-101(1999).
CC  -1- CATALYTIC ACTIVITY: DCTP + H(2)O -> DUTP + NH(3).
CC  -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: AP000059; BAA79288.1;
CC  InterPro: IPR003232; dCTP_deaminase.
CC  InterPro: IPR001428; dUTPase.
CC  Pfam: PF00652; dUTPase; 1.
CC  ProDom: PD004900; dCTP_deaminase; 1.
CC  HydroLase: Complete proteome.
CC  SEQUENCE 181 AA; 19894 MW; D8B6CBDC1722EFE9 CRC64;

Query Match          78.6%; Score 11; DB 1; Length 181;
Best Local Similarity 100.0%; Pred. No. 4e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  4 VDAGFEGOLTI 14
DB  117 VDAGFEGOLTI 127

RESULT 3
DCD_AQUAE          STANDARD;          PRT; 180 AA.
AC  O67539;
DT  20-AUG-2001 (Rel. 40, Created)
DT  20-AUG-2001 (Rel. 40, Last sequence update)
DE  20-AUG-2001 (Rel. 40, Last annotation update)
DE  DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE  DEAMINASE).
GN  DCD OR AQ.i607.
OS  Aquifex aeolicus.
OC  Bacteria; Aquificales; Aquificaceae; Aquifex.
OX  NCBI_TaxID=63363;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-VF5;
RX  MEDLINE=98196666; PubMed=9537320;
RA  Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA  Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA  Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT  "The complete genome of the hyperthermophilic bacterium Aquifex
RT  aeolicus.";
RL  Nature 392:353-358(1998).
CC  -1- CATALYTIC ACTIVITY: DCTP + H(2)O -> DUTP + NH(3).
CC  -1- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: AE000747; AAC07499.1;
CC  InterPro: IPR003232; dCTP_deaminase.
CC  InterPro: IPR001428; dUTPase.
CC  Pfam: PF00652; dUTPase; 1.
CC  ProDom: PD004900; dCTP_deaminase; 1.
CC  HydroLase: Complete proteome.
CC  SEQUENCE 180 AA; 20544 MW; B2710421A2FA48D6 CRC64;

Query Match          57.1%; Score 8; DB 1; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.061;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  4 VDAGFEGQ 11
DB  112 VDAGFEGQ 119

RESULT 4
VGH_BPS13          STANDARD;          PRT; 328 AA.
ID  VGH_BPS13;
AC  P07933;
DT  01-AUG-1988 (Rel. 08, Created)
DT  01-AUG-1988 (Rel. 08, Last sequence update)
DT  01-OCT-1996 (Rel. 34, Last annotation update)
DE  MINOR SPIKE PROTEIN (H PROTEIN) (PILOT PROTEIN).
GN  H.
OS  Bacteriophage S13.
OC  Viruses; ssDNA viruses; Microviridae; Microvirus.
OX  NCBI_TaxID=10844;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  Lau P.C.K., Spencer J.H.;
RT  "Nucleotide sequence and genome organization of bacteriophage S13
RT  DNA.";
RL  Gene 40:273-284(1985).
CC  -1- FUNCTION: MINOR SPIKE COMPONENT OF THE VIRAL SHELL. H PROTEIN IS
CC  INVOLVED IN THE EJECTION OF THE PHAGE DNA IN THE HOST AND IS
CC  INJECTED WITH THE DNA IN THE PERIPLASMIC SPACE OF THE HOST.
CC  -1- SUBUNIT: THE VIRION IS COMPOSED OF 60 COPIES EACH OF THE F, G, AND
CC  J PROTEINS AND 12 COPIES OF THE H PROTEIN. THERE ARE 12 SPIRES
CC  WHICH ARE EACH COMPOSED OF 5 G AND ONE H PROTEINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL: M14428; AAA32592.1;
CC  PIR: JS0459; JS0459.

```

```

KW Coat protein.
SQ SEQUENCE 328 AA; 34405 MW; EEABE8053765177F CRC64;

Query Match          50.0%; Score 7; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGFE 9
    |||||
DB 143 IVDAGFE 149

RESULT 5
T3RE_SALTY          STANDARD; PRT; 984 AA.
ID T3RE_SALTY
AC P40815;1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TYPE III RESTRICTION-MODIFICATION SYSTEM STYLT1 ENZYME RES
DE (EC 3.1.21.5).
GN RES.

OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT7;
RX MEDLINE=9325265; PubMed=8387444;
RA Dartois V., de Backer O., Colson C.;
RT "Sequence of the Salmonella typhimurium StyLT1
RT restriction-modification genes: homologies with EcoPI and EcoPI5
RT type-III R-M systems and presence of helicase domains.";
RL Gene 127:105-110(1993).
CC -1- FUNCTION: CLEAVES DNA SOME 25 BASE-PAIRS DOWNSTREAM FROM THE
CC RECOGNITION SITE. MAY ALSO ACT AS A HELICASE INVOLVED IN
CC UNWINDING DNA AT THE CLEAVAGE SITE. PROTEIN ONLY REQUIRED FOR
CC RESTRICTION BUT NEEDS THE PRESENCE OF THE MODIFICATION ENZYME.
CC -1- COFACTOR: MAGNESIUM AND ATP.
CC -1- SUBUNIT: CONTAINS TWO DIFFERENT SUBUNITS: RES AND MOD.
CC -1- SIMILARITY: WITH OTHER TYPE III RES PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M90544; -; NOT_ANNOTATED_CDS.
DR PIR; JN0658; JN0658.
DR REBASE; 1788; StyLT1.
KW Hydrolase; Nuclease; Endonuclease; Restriction system; Helicase.
SQ SEQUENCE 984 AA; 112970 MW; 8649B1AB5E9456BE CRC64;

Query Match          50.0%; Score 7; DB 1; Length 984;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDAGFE 10
    |||||
DB 315 VDAGFE 321

RESULT 6
ARYL_FELCA          STANDARD; PRT; 258 AA.
ID ARYL_FELCA
AC O62696;
DT 15-JUL-1999 (Rel. 38, Created)

15-JUL-1999 (Rel. 38, Last sequence update)
15-JUL-1999 (Rel. 38, Last annotation update)
DE ARYLAMINE N-ACETYLTRANSFERASE (EC 2.3.1.5) (ARYLAMINE ACETYLASE
DE (FRAGMENT)).
GN NAT.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99145352; PubMed=10022754;
RA Trepanier L.A., Cribb A.E., Spielberg S.P., Ray K.;
RT "Deficiency of cytosolic arylamine N-acetylation in the domestic cat
RT and wild felids caused by the presence of a single NAT1-like gene.";
RL Pharmacogenetics 8:169-179(1998).
CC -1- FUNCTION: PARTICIPATES IN THE DETOXIFICATION OF A PLETHORA OF
CC HYDRAZINE AND ARYLAMINE DRUGS.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE = COA +
CC N-ACETYLARYLAMINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF030398; AAC18940.1; -
DR InterPro; IPR001447; Acetyltransf2.
DR Pfam; PF00797; Acetyltransf2.1.
DR ProDom; PD002783; Acetyltransf2; 1.
KW Transferase; Acyltransferase.
FT NON_TER 1
FT NON_TER 258
SQ SEQUENCE 258 AA; 30138 MW; AD31B8745E79F2C3 CRC64;

Query Match          42.9%; Score 6; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
    |||||
DB 111 IVDAGF 116

RESULT 7
DUT_PRIVKA          STANDARD; PRT; 268 AA.
ID DUT_PRIVKA
AC Q90030; Q85226;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE (EC 3.6.1.23)
DE (DUTPASE) (DUTP PYROPHOSPHATASE).
GN UL50.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96135243; PubMed=8551588;
RA Joens A., Mettenleiter T.C.;
RT "Identification and characterization of pseudorabies virus dUTPase.";
RL J. Virol. 70:1242-1245(1996).
RN [2]
RP SEQUENCE OF 196-268 FROM N.A.
RX MEDLINE=95363968; PubMed=7637001;

```

BAUMEISTER J., Klupp B.G., Mettenleiter T.C.:
 "Pseudorabies virus and equine herpesvirus 1 share a nonessential
 gene which is absent in other herpesviruses and located adjacent to a
 highly conserved gene cluster.";
 J. Virol. 69:5560-5567(1995).
 CC -1- FUNCTION: THIS ENZYME IS INVOLVED IN NUCLEOTIDE METABOLISM: IT
 PRODUCES DUMP, THE IMMEDIATE PRECURSOR OF THYMIDINE NUCLEOTIDES
 AND IT DECREASES THE INTRACELLULAR CONCENTRATION OF DUTP SO THAT
 URACIL CANNOT BE INCORPORATED INTO DNA.
 CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
 CC -1- PATHWAY: DE NOVO SYNTHESIS OF THYMIDYLATE.
 CC -1- SIMILARITY: BELONGS TO THE DUTPASE FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U38547; AAB02855.1; -;
 DR HSP; P16088; IDUT.
 DR InterPro: IPR001428; dUTPase.
 DR Pfam; PF00692; dUTPase; 1.
 DR Hydrolase; Nucleotide metabolism.
 SQ SEQUENCE 268 AA; 28617 MW; 79D7CFA999204776 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 268;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAFG 8
 D 83 IVDAFG 88

RESULT 8
 ARYL_CHICK
 ID ARYL_CHICK STANDARD; PRT; 287 AA.
 AC P12275;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ARYLAMINE N-ACETYLTRANSFERASE, LIVER ISOZYME (EC 2.3.1.5) (ARYLAMINE
 ACETYLASE).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=88277945; PubMed=2897359;
 RA Ono S., Ohtomi M., Sakamoto Y., Ueyemura K., Deguchi T.;
 RT "Arylamine N-acetyltransferase from chicken liver II. Cloning of cDNA
 and expression in Chinese hamster ovary cells.";
 RL J. Biol. Chem. 263:7534-7538(1988).
 RN [2]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=88277945; PubMed=2897359;
 RA Deguchi T., Sakamoto Y., Sasaki Y., Ueyemura K.;
 RT "Arylamine N-acetyltransferase from chicken liver. I. Monoclonal
 antibodies, immunoaffinity purification, and amino acid sequences.";
 RL J. Biol. Chem. 263:7528-7533(1988).
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE = COA +
 N-ACETYLARYLAMINE.
 CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; J03737; AAA48590.1; -;
 DR PIR; A28168; A28168.
 DR PIR; A28167; A28167.
 DR InterPro: IPR001447; Acetyltransf2.
 DR Pfam; PF00797; Acetyltransf2; 1.
 DR ProDom; PD002783; Acetyltransf2; 1.
 DR Transf; Acetyltransferase; Multigene family.
 FT ACT_SITE 68 68 BY SIMILARITY
 SQ SEQUENCE 287 AA; 33915 MW; 6E2919AD2979E210 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 287;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAFG 8
 D 120 IVDAFG 125

RESULT 9
 ARYL_CHICK
 ID ARYL_CHICK STANDARD; PRT; 290 AA.
 AC P13913;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ARYLAMINE N-ACETYLTRANSFERASE, PINEAL GLAND ISOZYME NAT-10
 (EC 2.3.1.5) (ARYLAMINE ACETYLASE).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WHITE LEGHORN; TISSUE=Pineal gland;
 RX MEDLINE=90060108; PubMed=2583181;
 RA Ohtomi M., Sasaki M., Deguchi T.;
 RT "Two arylamine N-acetyltransferases from chicken pineal gland as
 identified by cDNA cloning.";
 RL Eur. J. Biochem. 195:253-261(1989).
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE = COA +
 N-ACETYLARYLAMINE.
 CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X16021; CAA34153.1; -;
 DR PIR; S06653; XYCHYO.
 DR InterPro: IPR001447; Acetyltransf2.
 DR Pfam; PF00797; Acetyltransf2; 1.
 DR ProDom; PD002783; Acetyltransf2; 1.
 DR Transf; Acetyltransferase; Multigene family.
 FT ACT_SITE 68 68 BY SIMILARITY
 SQ SEQUENCE 290 AA; 33925 MW; CBED93184E5F137C CRC64;

Query Match 42.9%; Score 6; DB 1; Length 290;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 IVDAGF 8
111111
Db 120 IVDAGF 125

RESULT 10

ARYL_HUMAN STANDARD; PRT; 290 AA.
AC P18440; O15300; O15159;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARYLAMINE N-ACETYLTRANSFERASE 1 (EC 2.3.1.5) (ARYLAMIDE ACETYLASE 1)
DE (ARYLAMINE N-ACETYLTRANSFERASE, MONOMORPHIC) (MNAT) (N-
DE ACETYLTRANSFERASE TYPE 1) (NAT-1).
GN NAT1 OR AAC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (NAT1*4).
RC TISSUE=Leukocyte;
RX MEDLINE=90253613; PubMed=2340091;
RA Blum M., Grant D.M., McBride W., Heim M., Meyer U.A.;
RT "Human arylamine N-acetyltransferase genes: isolation, chromosomal
RT localization, and functional expression.";
RL DNA Cell Biol. 9:193-203(1990).
RN [2]
RP SEQUENCE FROM N.A. (NAT1*5).
RC TISSUE=Liver;
RX MEDLINE=90170975; PubMed=1968463;
RA Ohsako S., Deguchi T.;
RT "Cloning and expression of cDNAs for polymorphic and monomorphic
RT arylamine N-acetyltransferases from human liver.";
RL J. Biol. Chem. 265:4630-4634(1990).
RN [3]
RP SEQUENCE FROM N.A. (NAT1*11).
RX MEDLINE=93183022; PubMed=8442668;
RA Vatsis K.P., Weber W.W.;
RT "Structural heterogeneity of Caucasian N-acetyltransferase at the
RT NAT1 gene locus.";
RL Arch. Biochem. Biophys. 301:71-76(1993).
RN [4]
RP SEQUENCE FROM N.A. (NAT1*11).
RC TISSUE=Blood;
RX MEDLINE=97312454; PubMed=9168895;
RA Doll M.A., Jiang W., Deitz A.C., Rustan T.D., Hein D.W.;
RT "Identification of a novel allele at the human NAT1 acetyltransferase
RT locus.";
RL Biochem. Biophys. Res. Commun. 233:584-591(1997).
RN [5]
RP SEQUENCE FROM N.A., AND VARIANTS NAT1*14 AND NAT1*17.
RX MEDLINE=98172173; PubMed=9511183;
RA Butcher N.J., Ilett K.F., Minchin R.F.;
RT "Functional polymorphism of the human arylamine N-acetyltransferase
RT type 1 gene caused by C190T and G560A mutations.";
RL Pharmacogenetics 8:67-72(1998).
RN [6]
RP SEQUENCE OF 1-22 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91282773; PubMed=1676262;
RA Ebisawa T., Deguchi T.;
RT "Structure and restriction fragment length polymorphism of genes for
RT human liver arylamine N-acetyltransferases.";
RL Biochem. Biophys. Res. Commun. 177:1252-1257(1991).
RN [7]
RP MUTAGENESIS OF ARG-64.
RX MEDLINE=97234849; PubMed=9173883;
RA Delomenie C., Goodfellow G.H., Krishnamoorthy R., Grant D.M.,
RA Dupret J.M.;

*Study of the role of the highly conserved residues Arg9 and Arg64 in
the catalytic function of human N-acetyltransferases NAT1 and NAT2 by
site-directed mutagenesis.";
RL Biochem. J. 323:207-215(1997).
RN [8]
RP VARIANT NAT1*14.
RX MEDLINE=98172172; PubMed=9511182;
RA Hughes N.C., Janezic S.A., McQueen K.L., Jewett M.A., Castranio T.,
RA Bell D.A., Grant D.M.;
RT "Identification and characterization of variant alleles of human
RT acetyltransferase NAT1 with defective function using
RT p-aminosalicylate as an in-vivo and in-vitro probe.";
RN Pharmacogenetics 8:55-66(1998).
RN [9]
RP VARIANTS NAT1*17; NAT1*21; NAT1*22; NAT1*24 AND NAT1*25.
RX MEDLINE=98347194; PubMed=9682272;
RA Lin H.J., Probst-Hensch N.M., Hughes N.C., Sakamoto G.T., Louie A.D.,
RA Kau I.H., Lin B.K., Lee D.B., Lin J., Frankl H.D., Lee E.R., Hardy S.,
RA Grant D.M., Haile R.W.;
RT "Variants of N-acetyltransferase NAT1 and a case-control study of
RT colorectal adenomas.";
RL Pharmacogenetics 8:269-281(1998).
RN [10]
RP FUNCTION: PARTICIPATES IN THE DETOXIFICATION OF A PLETHORA OF
HYDRAZINE AND ARYLAMINE DRUGS. CATALYZES THE N- OR O-ACETYLATION
OF VARIOUS ARYLAMINE AND HETEROCYCLIC AMINE SUBSTRATES AND IS ABLE
TO BIOACTIVATE SEVERAL KNOWN CARCINOGENS.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE = COA +
N-ACETYLARYLAMINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- POLYMORPHISM: N-ACETYLATION POLYMORPHISM IS DETERMINED BY A LOW OR
HIGH NAT ACTIVITY IN LIVER, IT HAS BEEN IMPLICATED IN THE ACTION
AND TOXICITY OF AMINE-CONTAINING DRUGS. SLOW ACETYLATION GENOTYPES
HAVE BEEN ASSOCIATED WITH SIGNIFICANT LUNG CANCER RISK. CANDIDATE
RISK FACTOR FOR SUSCEPTIBILITY TO NEURAL TUBE DEFECTS. THE NAT1*10
ALLELE HAS BEEN ASSOCIATED WITH INCREASED RISK OF COLON AND
URINARY BLADDER CANCERS AND WITH HIGHER LEVELS OF N-
ACETYLTRANSFERASE ACTIVITY AND DNA ADDUCTS IN AROMATIC AMINE TUMOR
TARGET ORGANS SUCH AS COLON AND URINARY BLADDER.
CC -1- MISCELLANEOUS: NAT1 WAS HISTORICALLY CONSIDERED TO BE MONOMORPHIC
IN NATURE BUT REPORTS OF ALLELIC VARIATIONS AT THE NAT1 LOCUS
SUGGEST THAT IT IS A POLYMORPHICALLY EXPRESSED ENZYME.
CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
CC -1- CAUTION: THE ALLELIC VARIATION ILE-149 DESIGNATED AS NAT1*17 IS
PART OF THE NAT1*11 ALLELIC VARIATION AS REPORTED BY THE
NOMENCLATURE COMMITTEE.
CC -1- DATABASE: NAME-NAT; NOTE-NAT alleles;
WWW="http://www.louisville.edu/medschool/pharmacology/NAT.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X17059; CAA34905.1; -;
DR EMBL; D90041; BAA14095.1; -;
DR EMBL; AF032677; AAB86878.1; -;
DR EMBL; AF032678; AAB86879.1; -;
DR EMBL; U80835; AAB62398.1; -;
DR EMBL; AF008204; AAB84384.1; -;
DR EMBL; M75164; AAA59905.1; -;
DR PIR; A34585; A34585.
DR PIR; B35729; B35729.
DR MIM; 108345; -;
DR InterPro; IPR001447; Acetyltransf2.
DR Pfam; PF00797; Acetyltransf2; 1.
DR ProDom; PD002783; Acetyltransf2; 1.
DR Transfaser; Acetyltransferase; Multigene family; Polymorphism.
FT ACT_SITE 68 68 BY SIMILARITY.
FT VARIANT 64 64 R -> W (IN NAT1*17; A SLOW ACETYLATOR;
HAS DEFECTIVE ENZYME ACTIVITY).
FT

```

FT FT /FTID-VAR_004606.
FT FT R -> T (IN NAT1*5).
FT FT /FTID-VAR_009510.
FT FT V -> I (IN NAT1*11; CATALYZES THE
FT FT N-ACETYLATION OF AROMATIC AMINES AND THE
FT FT O- AND N,O- ACETYLATION OF THEIR N-
FT FT HYDROXYLATED METABOLITES AT RATES UP TO
FT FT 2-FOLD HIGHER).
FT FT /FTID-VAR_004607.
FT FT RE -> TQ (IN NAT1*5).
FT FT /FTID-VAR_009511.
FT FT R -> Q (IN NAT1*14; A SLOW ACETYLATOR).
FT FT /FTID-VAR_009069.
FT FT M -> V (IN NAT1*21).
FT FT /FTID-VAR_009070.
FT FT S -> A (IN NAT1*11).
FT FT /FTID-VAR_009071.
FT FT D -> V (IN NAT1*22).
FT FT /FTID-VAR_009072.
FT FT E -> K (IN NAT1*24).
FT FT /FTID-VAR_009073.
FT FT I -> V (IN NAT1*25).
FT FT /FTID-VAR_009074.
FT FT R -> A M.O.K.; REDUCED ENZYMIC ACTIVITY.
FT FT MUTAGEN 64
FT FT SEQUENCE 290 AA; 33898 MW; C015F7F3D4830107 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
DB 120 IVDAGF 125

RESULT 11
ARYL_RABIT STANDARD; PRT; 290 AA.
AC P18605;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARYLAMINE N-ACETYLTRANSFERASE 1 (EC 2.3.1.5) (ARYLAMIDE ACETYLASE 1)
DE (ARYLAMINE N-ACETYLTRANSFERASE, MONOMORPHIC) (MNAT) (N-
DE ACETYLTRANSFERASE TYPE 1) (NAT-1).
GN NAT1 OR AAC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX MEDLINE=90384838; PubMed=2402454;
RA Blum N., Heim M., Meyer U.A.;
RT "Nucleotide sequence of rabbit NAT1 encoding monomorph N-
RL acetyltransferase."
RN Nucleic Acids Res. 18:5287-5287(1990).
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91302353; PubMed=2071601;
RA Sasaki Y., Ohsaka S., Deguchi T.;
RT "Molecular and genetic analyses of arylamine N-acetyltransferase
RT polymorphism of rabbit liver."
RL J. Biol. Chem. 266:13243-13250(1991).
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE - COA +
CC N-ACETYLARYLAMINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10108; BAA00989.1; -
DR EMBL; X53765; CAA37785.1; -
DR PIR; C39870; XYRBM.
DR PIR; S11220; S11220.
DR InterPro; IPR001447; Acetyltransf2.
DR Pfam; PF00797; Acetyltransf2; 1.
DR ProDom; PD002783; Acetyltransf2; 1.
KW Transferase; Acyltransferase; Multigene family.
FT ACT_SITE 68 BY SIMILARITY.
FT CONFLICT 285 N -> H (IN REF. 2).
SQ SEQUENCE 290 AA; 33780 MW; 31175E83651EE693 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
DB 120 IVDAGF 125

RESULT 12
ARY2_MESAU STANDARD; PRT; 290 AA.
ID ARY2_MESAU
AC P50293;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARYLAMINE N-ACETYLTRANSFERASE 2 (EC 2.3.1.5) (ARYLAMIDE ACETYLASE 2)
DE (ARYLAMINE N-ACETYLTRANSFERASE, POLYMORPHIC) (PNAT) (N-
DE ACETYLTRANSFERASE TYPE 2) (NAT-2) (AT-2).
GN NAT2 OR AAC2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=94193006; PubMed=8144033;
RA Farquison R.J., Doll M.A., Baumstark B.R., Hein D.W.;
RT "Polymorphic arylamine N-acetyltransferase encoding gene (NAT2) from
RL homozygous, rapid and slow acetylators congenic Syrian hamsters."
RN Gene 140:247-249(1994).
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94332979; PubMed=8055637;
RA Land S.J., Jones R.F., King C.M.;
RT "Biochemical and genetic analysis of two acetyltransferases from
RL hamster tissues that can metabolize aromatic amine derivatives."
RN Carcinogenesis 15:1585-1595(1994).
RP SEQUENCE FROM N.A.
RX MEDLINE=94362736; PubMed=7915937;
RA Nagata K., Ozawa S., Miyata M., Shimada M., Yamazoe Y., Kato R.;
RT "Primary structure and molecular basis of polymorphic appearance of
RL an acetyltransferase (AT-II)* in hamsters."
CC Pharmacogenetics 4:91-100(1994).
CC -1- FUNCTION: PARTICIPATES IN THE DETOXIFICATION OF A PLETHORA OF
CC HYDRAZINE AND ARYLAMINE DRUGS.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE - COA +
CC N-ACETYLARYLAMINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

```

```

CC -1- POLYMORPHISM: THERE ARE TWO FORMS OF NAT2: A RAPID ISOFORM AND A
CC SLOW ISOFORM.
CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U03468; AAB60524.1; -
CC EMBL; U03467; AAB60523.1; -
CC EMBL; L24912; AAA21829.1; -
CC EMBL; S72005; AAB31917.1; -
CC EMBL; S72007; AAB31918.1; -
CC InterPro: IPR001447; Acetyltransf2.
CC Pfam; PF00797; Acetyltransf2; 1.
CC ProDom; PD002783; Acetyltransf2; 1.
CC Transferase; Acyltransferase; Multigene family; Polymorphism.
KW ACT_SITE 68 BY SIMILARITY.
FT VARIANT 243 290 MISSING (IN SLOW ISOFORM).
FT SEQUENCE 290 AA; 33832 MW; 54C10F5F6990F6FD CRC64;
SQ
Query Match 42.9%; Score 6; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 IVDAGF 8
Db 120 IVDAGF 125
RESULT 13
ARY2_MOUSE
ID ARY2_MOUSE STANDARD; PRT; 290 AA.
AC P50295;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARYLAMINE N-ACETYLTRANSFERASE 2 (EC 2.3.1.5) (ARYLAMIDE ACETYLASE 2)
DE (N-ACETYLTRANSFERASE TYPE 2) (NAT-2).
GN NAT2 OR AAC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 AND A/J;
RX MEDLINE=91342604; PubMed=1875909;
RA Martell K.J., Vatsis K.P., Weber W.W.;
RT "Molecular genetic basis of rapid and slow acetylation in mice.";
RL Mol. Pharmacol. 40:218-227(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HEJ; TISSUE=Heart;
RA Hein D.W., Doll M.A.;
RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RC STRAIN=BA1B/C; TISSUE=Liver;
RX MEDLINE=94379961; PubMed=7545952;
RA Kelly S.L., Sim E.;
RT "Arylamine N-acetyltransferase in Balb/c mice: identification of a
RT novel mouse isoenzyme by cloning and expression in vitro.";
RL Biochem. J. 302:347-353(1994).
RN [4]
RP CHARACTERIZATION.
RC
RX MEDLINE=92382565; PubMed=1513324;
RA Martell K.J., Levy G.N., Weber W.W.;

```

```

RT "Cloned mouse N-acetyltransferases: enzymatic properties of expressed
RT Nat-1 and Nat-2 gene products.";
RL Mol. Pharmacol. 42:265-272(1992).
CC -1- FUNCTION: PARTICIPATES IN THE DETOXIFICATION OF A PLETHORA OF
CC HYDRAZINE AND ARYLAMINE DRUGS. 2-AMINOFLUORENE AND P-AMINOBENZAZOIC
CC ACID (PABA) ARE PREFERRED SUBSTRATES FOR NAT-2. LESS ACTIVITY WITH
CC ANISIDINE AND BARELY DETECTABLE WITH SMZ.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE = COA +
CC N-ACETYLARYLAMINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- POLYMORPHISM: THERE ARE TWO FORMS OF NAT2: A RAPID/STABLE ISOFORM
CC (ASN-99) AND A SLOW/UNSTABLE ISOFORM (ILE-99).
CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U35886; AAA78943.1; -
CC EMBL; U35887; AAA78944.1; -
CC EMBL; U37249; AAA80353.1; -
CC EMBL; U37250; AAA80354.1; -
CC MGD; MGI:109201; Nat2.
CC InterPro: IPR001447; Acetyltransf2.
CC Pfam; PF00797; Acetyltransf2; 1.
CC ProDom; PD002783; Acetyltransf2; 1.
CC Transferase; Acyltransferase; Multigene family; Polymorphism.
KW ACT_SITE 68 BY SIMILARITY.
FT VARIANT 99 99 N->I (IN NAT2*9; SLOW/UNSTABLE
FT SEQUENCE 290 AA; 33701 MW; 704E000DE48CE557 CRC64;
SQ
Query Match 42.9%; Score 6; DB 1; Length 290;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 IVDAGF 8
Db 120 IVDAGF 125
RESULT 14
ARY2_RABIT
ID ARY2_RABIT STANDARD; PRT; 290 AA.
AC P11246;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARYLAMINE N-ACETYLTRANSFERASE 2 (EC 2.3.1.5) (ARYLAMIDE ACETYLASE 2)
DE (ARYLAMINE N-ACETYLTRANSFERASE, POLYMORPHIC) (PNAT) (N-
DE ACETYLTRANSFERASE TYPE 2) (NAT-2).
GN NAT2 OR AAC2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX MEDLINE=89263801; PubMed=2726493;
RA Blum M., Grant D.M., Demierre A., Meyer U.A.;
RT "Nucleotide sequence of a full-length cDNA for arylamine N-
RT acetyltransferase from rabbit liver.";
RL Nucleic Acids Res. 17:3589-3589(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC
RX MEDLINE=90384846; PubMed=2402461;
RA Blum M., Heim M., Meyer U.A.;

```

```
RT "Nucleotide sequence of rabbit NAT2 encoding polymorphic liver
RT arylamine N-acetyltransferase (NAT2).";
RL Nucleic Acids Res. 18:5295-5295(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=91302353; PubMed=2071601;
RA Sasaki Y., Ohsako S., Deguchi T.;
RT Molecular and genetic analyses of arylamine N-acetyltransferase
RT polymorphism of rabbit liver.;
RL J. Biol. Chem. 266:13243-13250(1991).
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE - COA +
CC N-ACETYLARYLAMINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10109; BAA00990.1; -
DR EMBL; D10111; BAA00991.1; -
DR EMBL; X14673; CAA32803.1; -
DR EMBL; X53767; CAA37786.1; -
DR PIR; A39870; XYRBPB.
DR InterPro; IPR001447; Acetyltransf2.
DR Pfam; PF00797; Acetyltransf2; 1.
DR ProDom; PD002783; Acetyltransf2; 1.
DR Transferrase; Acyltransferase; Multigene family.
KW ACT_SITE 68 68 BY SIMILARITY.
FT VARIANT 121 121 V -> I (IN NAT2*21, SLOW ACETYLATOR).
FT VARIANT 266 266 V -> I (IN NAT2*21, SLOW ACETYLATOR).
SQ SEQUENCE 290 AA; 33649 MW; CF0ECBED398C352A CRC64;

Query Match 42.9%; Score 6; DB 1; Length 290;
Best Local Similarity 100.0%; Pred.No.12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
DB 120 IVDAGF 125

RESULT 15
ARY2_RAT ID ARY2_RAT STANDARD; PRT; 290 AA.
AC P50298;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ARYLAMINE N-ACETYLTRANSFERASE 2 (EC 2.3.1.5) (ARYLAMIDE ACETYLASE 2)
DE (N-ACETYLTRANSFERASE TYPE 2) (NAT-2) (AT-2).
GN NAT2 OR AAC2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Pineal gland;
RX MEDLINE=95188895; PubMed=7882993;
RA Ebisawa T., Sasaki Y., Deguchi T.;
RT Complementary DNAs for two arylamine N-acetyltransferases with
RT identical 5' non-coding regions from rat pineal gland.;
RL Eur. J. Biochem. 228:129-137(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WKY/CRLBR; TISSUE=Heart;
RX MEDLINE=96034572; PubMed=8528272;
RA Doi M.A., Hein D.W.;
```

```
RT "Cloning, sequencing and expression of NAT1 and NAT2 encoding genes
RT from rapid and slow acetylator inbred rats.";
RL Pharmacogenetics 5:247-251(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RA Jones R.F., Gott B., Land S.J., Park J., King C.M.;
RL Submitted (XXX-1993), to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PARTICIPATES IN THE DETOXIFICATION OF A PLETHORA OF
CC HYDRAZINE AND ARYLAMINE DRUGS. ACETYLATES ONLY ARYLAMINES.
CC -1- CATALYTIC ACTIVITY: ACETYL-COA + ARYLAMINE - COA +
CC N-ACETYLARYLAMINE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- POLYMORPHISM: THERE ARE TWO FORMS OF NAT2: A RAPID ISOFORM
CC (NAT2*21A) AND A SLOW ISOFORM (NAT2*21B).
CC -1- SIMILARITY: BELONGS TO THE ARYLAMINE N-ACETYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U01348; AAA70161.1; -
DR EMBL; U23418; AAB53956.1; -
DR EMBL; U19272; AAB60501.1; -
DR EMBL; U17261; AAA56772.1; -
DR InterPro; IPR001447; Acetyltransf2.
DR Pfam; PF00797; Acetyltransf2; 1.
DR ProDom; PD002783; Acetyltransf2; 1.
DR Transferrase; Acyltransferase; Multigene family; Polymorphism.
KW ACT_SITE 68 68 BY SIMILARITY.
FT VARIANT 121 121 V -> I (IN NAT2*21, SLOW ACETYLATOR).
FT VARIANT 266 266 V -> I (IN NAT2*21, SLOW ACETYLATOR).
SQ SEQUENCE 290 AA; 33756 MW; A02BF839A230F84D CRC64;

Query Match 42.9%; Score 6; DB 1; Length 290;
Best Local Similarity 100.0%; Pred.No.12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
DB 120 IVDAGF 125
```

Search completed: January 31, 2002, 13:39:21
Job time: 81 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:11 ; Search time 78.64 Seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-74
Perfect score: 14
Sequence: 1 AGWIDAGFKGKITL 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	204	2 F64353	dCTP deaminase (EC
2	7	50.0	315	1 D35156	thioredoxin reduct
3	7	50.0	443	1 QCBTPT	hypothetical prote
4	7	50.0	564	2 T26053	hypothetical prote
5	6	42.9	88	2 G81178	conserved hypotet
6	6	42.9	88	2 B81926	hypothetical prote
7	6	42.9	103	2 B64334	hypothetical prote
8	6	42.9	183	1 I64006	hypothetical prote
9	6	42.9	213	2 T17241	hypothetical prote
10	6	42.9	233	2 S74064	hypothetical prote
11	6	42.9	287	1 S72569	probable aldehyde
12	6	42.9	296	2 T48401	histone deacetylase
13	6	42.9	311	2 B70451	prephenate dehydro
14	6	42.9	329	2 S04132	photosystem II oxy
15	6	42.9	352	2 H70338	threonine synthase
16	6	42.9	356	2 E64048	nucleotide-binding
17	6	42.9	356	2 T10888	iron utilization p
18	6	42.9	358	2 D81345	cellulose periplasm
19	6	42.9	444	2 T46602	cellobiose/cellotr
20	6	42.9	455	2 D81809	hypothetical prote
21	6	42.9	485	2 T18793	hypothetical prote
22	6	42.9	508	1 A30007	dolichyl-diphospho
23	6	42.9	508	1 ISRTSS	protein disulfide-
24	6	42.9	509	1 ISMSSS	protein disulfide-
25	6	42.9	509	1 A38362	protein disulfide-
26	6	42.9	525	2 T47409	hypothetical prote
27	6	42.9	575	2 T38406	probable flavoprot
28	6	42.9	731	2 T44752	probable malate sy
29	6	42.9	883	2 T37208	hypothetical prote

30	6	42.9	1061	2 T18085	DNA topoisomerase
31	6	42.9	1132	1 QSBPL	host specificity p
32	6	42.9	1138	2 D85584	probable tail comp
33	6	42.9	1160	2 A46423	transcription fact
34	6	42.9	1849	2 T14096	guanine nucleotide
35	6	42.9	2104	2 H86127	hypothetical prote
36	5	35.7	19	2 A60894	gamma-crystallin I
37	5	35.7	47	2 S69146	gamma-thionin Si-a
38	5	35.7	47	2 S11529	gamma-purothionin
39	5	35.7	47	2 S11530	purothionin gamma
40	5	35.7	47	2 S13849	hordothionin gamma
41	5	35.7	47	2 A58319	gamma-zeathionin I
42	5	35.7	47	2 S69144	gamma-thionin Si-a
43	5	35.7	48	2 S13963	alpha-amylase inh
44	5	35.7	53	2 I60384	gene T1 protein -
45	5	35.7	92	2 I39925	transcription regu

ALIGNMENTS

RESULT 1
F64353
dCTP deaminase (EC 3.5.4.13) MJ0430 [similarity] - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: F64353
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A;Reference number: A64300; MUID:96337999
A;Accession: F64353
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-204 <BUL>
A;Cross-references: GB:U67494; GB:L77117; NID:gl591120; PIDN:ANB98415.1; PID:gl591133
C;Genetics:
A;Map position: REV386963-386349
C;Superfamily: dCTP deaminase
C;Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGWIDAGFKGKITL 14
|||||
DB 131 AGWIDAGFKGKITL 144

RESULT 2
D35156
thioredoxin reductase (NADPH) (EC 1.6.4.5) - Eubacterium acidaminophilum
N;Alternate names: dihydrolipoamide dehydrogenase [misidentification]
C;Species: Eubacterium acidaminophilum
C;Date: 06-Nov-1998 #sequence_revision 06-Nov-1998 #text_change 04-Feb-2000
C;Accession: S38988; D35156
R;Luebbbers, M.; Andreesen, J.R.
Eur. J. Biochem. 217, 791-798, 1993
A;Title: Components of glycine reductase from Eubacterium acidaminophilum. Cloning, s
A;Reference number: S38988; MUID:94039119
A;Accession: S38988
A;Molecule type: DNA
A;Residues: 1-315 <LUE>
A;Cross-references: GB:I04500; NID:g2708733
R;Dietrichs, D.; Meyer, M.; Schmidt, B.; Andreesen, J.R.
J. Bacteriol. 172, 2088-2095, 1990
A;Title: Purification of NADPH-dependent electron-transferring flavoproteins and N-te
A;Reference number: A35156; MUID:90202731

A:Accession: D35156
A:Molecule type: Protein
A:Residues: 1-33, X'35-46, X'48, D', 50-55 <DIE>
A:Note: The designation "atypically small dihydrolipoamide dehydrogenase" was revised in
C:Genetics:
A:Gene: trxB
A:Start codon: GTG
A:Complex: homodimer; each chain has a tightly associated but noncovalently bound FAD
C:Function:
A:Description: catalyzes the reversible reduction of oxidized thioredoxin by NADPH
C:Superfamily: Thioredoxin reductase; thioredoxin reductase homology
C:Keywords: FAD; flavoprotein; homodimer; NADP; oxidoreductase; redox-active disulfide
F:1-309/Domain: thioredoxin reductase homology <TRXB>
F:6-34/Region: beta-alpha-beta FAD nucleotide-binding fold
F:146-173/Region: beta-alpha-beta NADP nucleotide-binding fold
F:134-137/Disulfide bonds: redox-active #staple predicted

Query Match 50.0% Score 7; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 FKGRITL 14
| | | | |
Db 252 FKGRITL 258

RESULT 3
QOBYPT
hypothetical protein YER152c - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 30-Jun-1991 #sequence-revision 12-May-1995 #text_change 23-Mar-2001
C:Accession: S50655; C36328; S05725; S12290
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lambda
A:Reference number: S50430
A:Accession: NS0655
A:Molecule type: DNA
A:Residues: 1-443 <DIE>
A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64679.1; PID:g603392; GSPDB:GN0000
R:Ohmen, J.D.; Burke, K.A.; McEwen, J.E.
Mol. Cell. Biol. 10, 3027-3035, 1990
A:Title: Divergent overlapping transcripts at the PET122 locus in Saccharomyces cerevisiae
A:Reference number: A36328; MUID:90258894
A:Accession: C36328
A:Molecule type: DNA
A:Residues: 1-230 <OHM>
A:Cross-references: GB:X07558; NID:g4127; PIDN:CAA30440.1; PID:g4129
R:Ohmen, J.D.; Kloeckner-Gruissem, B.; McEwen, J.E.
Nucleic Acids Res. 16, 10783-10802, 1988
A:Title: Molecular cloning and nucleotide sequence of the nuclear PET122 gene required for
A:Reference number: S05724; MUID:89083497
A:Accession: S05725
A:Molecule type: DNA
A:Residues: 1-56, 'CGA', 60-61, 63-66, 'V', 68-72, 'T', 74-76, 'R', 78-86, 'V', 88-122, 'SDQQ', 128-1
A:Cross-references: EMBL:X07558
A:Note: this sequence has been revised in reference A36328
C:Genetics:
A:Gene: MIPS:YER152C
A:Map position: 5R
C:Superfamily: yeast hypothetical protein YER152c

Query Match 50.0% Score 7; DB 1; Length 443;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 DAGFKGK 11
| | | | |
Db 131 DAGFKGK 137

RESULT 4
T26053
hypothetical protein W01C9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26053
R:Sims, M.
submitted to the EMBL Data Library, June 1995
A:Reference number: Z20143
A:Accession: T26053
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-564 <WIL>
A:Cross-references: EMBL:Z49969; PIDN:CAA90270.1; GSPDB:GN00020; CESP:W01C9.2
A:Experimental source: clone W01C9
C:Genetics:
A:Gene: CESP:W01C9.2
A:Map position: 2
A:Introns: 30/1; 64/1; 100/3; 124/3; 229/3; 272/3; 536/2

Query Match 50.0% Score 7; DB 2; Length 564;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 FKGRITL 14
| | | | |
Db 2 FKGRITL 8

RESULT 5
G81178
conserved hypothetical protein NMB0606 [imported] - Neisseria meningitidis (strain MC
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81178
R:Rettlein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Hart, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Oid, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: G81178
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-88 <TET>
A:Cross-references: GB:AE002416; GB:AE002098; NID:g725832; PIDN:AAF41033.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0606

Query Match 42.9% Score 6; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 AGFKGK 11
| | | | |
Db 50 AGFKGK 55

RESULT 6
B81926
hypothetical protein NMA0811 [imported] - Neisseria meningitidis (strain Z2491 serogr
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence-revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: B81926
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: A81775; MUID:20225556

A:Accession: B81926
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-88 <PAR>
 A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84093.1; PID:g737953
 A:Experimental source: serogroup A, strain 22491
 C:Genetics:
 A:Gene: NMA0811

Query Match 42.9%; Score 6; DB 2; Length 88;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11

Db 50 AGFKGK 55

RESULT 7

B64334

hypothetical protein MJ0273 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: B64334

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Bult, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Olson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A:Reference number: A64300; MUID:96337999

A:Accession: B64334

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-103 <BUL>

A:Cross-references: GB:U67482; GB:L77117; NID:g2826267; PIDN:AAB98261.1; PID:g1590998; T

C:Genetics:

A:Map position: FOR258104-258415

A:Start codon: TTG

Query Match

Best Local Similarity 42.9%; Score 6; DB 2; Length 103;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GWIDAG 7

Db 79 GWIDAG 84

RESULT 8

I64006

hypothetical protein HI0389 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000

C:Accession: I64006

R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A:Title: whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630

A:Accession: I64006

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-183 <TIGR>

A:Cross-references: GB:U32722; GB:L42023; NID:g1573348; PIDN:AAC22047.1; PID:g1573359; T

C:Superfamily: rnd protein

Query Match

42.9%; Score 6; DB 1; Length 183;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KGKITL 14

Db 2 KGKITL 7

RESULT 9

T17241

hypothetical protein DKFZp434L057.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T17241

R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18722

A:Accession: T17241

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-213 <KOE>

A:Cross-references: EMBL:AL117446

A:Experimental source: adult testis; clone DKFZp434L057

C:Genetics:

A:Note: DKFZp434L057.1

Query Match

Best Local Similarity 42.9%; Score 6; DB 2; Length 213;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11

Db 73 AGFKGK 78

RESULT 10

S74064

hypothetical protein c0133 - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C>Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000

C:Accession: S74064

R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, M.L. Microbiol. 22, 175-191, 1996

A:Title: Organizational characteristics and information content of an archaeal genome

A:Reference number: S73076; MUID:97055432

A:Accession: S74064

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-233 <SEN>

A:Cross-references: EMBL:Y08256; NID:g1707679; PID:g1707763

A:Experimental source: strain P2

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 199

C:Genetics:

A:Start codon: TTG

C:Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MT

Query Match

Best Local Similarity 42.9%; Score 6; DB 2; Length 233;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFKGKI 12

Db 164 GFKGKI 169

RESULT 11

S72569

probable aldehyde reductase (EC 1.1.1.-) C35D10.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S72569

R;Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burton, L.; L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Johnston, L.; Jones, B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifkin, L.; Roopra, A.; Saunders, D. submitted to the EMBL Data Library, February 1995
 A:Authors: Shownkeen, R.; Smalton, N.; Smith, A.; Sonhammer, E.; Staden, R.; Sulston, J. proat, J.; Wohlman, P.
 A:Description: The C. elegans genome project: Contiguous nucleotide sequence of over two
 A:Reference number: S72566
 A:Accession: S72569
 A:Molecule type: DNA
 A:Residues: 1-287 <WIL>
 A:Cross-references: EMBL:U21324; NID:g687879; PIDN:AAA62562.1; PID:g687885
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Map position: 3
 A:Introns: 20/3; 69/2; 195/3; 261/3
 A:Note: C35D10.6
 C:Superfamily: aldehyde reductase
 C:Keywords: oxidoreductase

Query Match 42.9%; Score 6; DB 1; Length 287;
 Best Local Similarity 100.0%; Pred. No. 29; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0;

Qy 4 IDAGFK 9
 Db 31 IDAGFK 36

RESULT 12

T48401
 histone deacetylase-like protein - Arabidopsis thaliana
 N:Alternate names: protein F17C15.160
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T48401
 R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; submitted to the Protein Sequence Database, March 2000
 A:Reference number: 224492
 A:Accession: T48401
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-296 <BEV>
 A:Cross-references: EMBL:AL162506
 A:Experimental source: cultivar Columbia; BAC clone F17C15
 C:Genetics:
 A:Map position: 5
 A:Introns: 5/1; 28/3; 93/3; 116/3; 164/3; 195/3; 227/1
 A:Note: F17C15.160

Query Match 42.9%; Score 6; DB 2; Length 296;
 Best Local Similarity 100.0%; Pred. No. 30; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0;

Qy 6 AGFRGK 11
 Db 105 AGFRGK 110

RESULT 13

B70451
 prephenate dehydrogenase - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 19-May-2000
 C:Accession: B70451
 R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666
 A:Accession: B70451

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-311 <AQF>
 A:Cross-references: GB:AE000754; NID:g2984047; PIDN:AA07589.1; PID:g2984048; GB:AE00
 A:Experimental source: strain VF5
 C:Genetics:
 A:Gene: tyrA
 C:Superfamily: cyclohexadienyl dehydrogenase

Query Match 42.9%; Score 6; DB 2; Length 311;
 Best Local Similarity 100.0%; Pred. No. 31; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0;

Qy 7 GFKGKI 12
 Db 53 GFKGKI 58

RESULT 14

S04132
 Photosystem II oxygen-evolving complex protein 1 precursor - garden pea
 N:Alternate names: extrinsic water-oxidizing complex 33k chain; Photosystem II extrin
 C:Species: Pisum sativum (garden pea)
 C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 11-Jan-2000
 C:Accession: S04132; A30457; S03269; S28886
 R:Wales, R.; Newman, B.J.; Pappin, D.; Gray, J.C.
 Plant Mol. Biol. 12, 439-451, 1989
 A:Title: The extrinsic 33 kDa polypeptide of the oxygen-evolving complex of photosyst
 A:Reference number: S04132
 A:Accession: S04132
 A:Molecule type: mRNA
 A:Residues: 1-329 <WAL>
 A:Cross-references: EMBL:X15350; NID:g20620; PID:g20621
 A:Accession: A30457

A:Molecule type: protein
 A:Residues: 82-88; 'X', 90-94; 'X', 96; 'X', 98-101 <WAL>
 R:Murata, N.; Kajihara, H.; Fujimura, Y.; Miyao, M.; Murata, T.; Watanabe, A.; Shinoza
 Prog. Photosyn. Res. 1, 701-704, 1987
 A:Title: Partial amino acid sequences of the proteins of pea and spinach photosystem
 A:Reference number: S03269
 A:Accession: S03269
 A:Molecule type: protein
 A:Residues: 82-103; 'XXX', 107-109; 'X', 111 <MUR>
 C:Superfamily: Photosystem II oxygen-evolving complex protein 1
 C:Keywords: chloroplast; membrane-associated complex; peripheral membrane protein; ph
 F:1-81/Domain: transit peptide (chloroplast) #status predicted <NP>
 F:82-329/Product: Photosystem II oxygen-evolving complex protein 1 #status predicted

Query Match 42.9%; Score 6; DB 2; Length 329;
 Best Local Similarity 100.0%; Pred. No. 32; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 0;

Qy 9 KGKITL 14
 Db 278 KGKITL 283

RESULT 15

H70338
 threonine synthase (EC 4.2.99.2) - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 22-Jun-1999
 C:Accession: H70338
 R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666
 A:Accession: H70338
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A;Residues: 1-352 <AQF>
A;Cross-references: GB:AE000689; NID:g2983082; PIDN:AAC06690.1; PID:g2983084; GB:AE000689
A;Experimental source: strain VF5
C;Genetics:
A;Gene: thrC2
C;Function:
A;Description: catalyzes hydrolyzation of O-phospho-L-homoserine to L-threonine and orthophosphate
A;Pathway: threonine biosynthesis
A;Note: pyridoxal phosphate cofactor
C;Superfamily: threonine dehydratase
C;Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate; threonine biosynthesis
F:63/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 42.9%; Score 6; DB 2; Length 352;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFKGI 12
|||||
Db 44 GFKGI 49

Search completed: January 31, 2002, 13:20:12
Job time: 113 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:12 ; Search time 78.64 seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-75

Perfect score: 14

Sequence: 1 SAVHDPGVEGRPEY 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	161	2 E64437	probable dCTP deam
2	7	50.0	1197	1 A48350	DNA-binding protei
3	6	42.9	259	2 T29427	probable exonuclea
4	6	42.9	634	2 E83263	hypothetical prote
5	6	42.9	1025	2 S34839	collagen alpha 1(V
6	5	35.7	69	2 T04135	auxin-inducible pr
7	5	35.7	88	1 EDBE51	immediate-early-5
8	5	35.7	119	2 A82767	50S ribosomal prot
9	5	35.7	136	2 T03008	probable regulator
10	5	35.7	140	2 S74547	hypothetical prote
11	5	35.7	140	2 T36136	hypothetical prote
12	5	35.7	147	2 I51014	MHC class I protei
13	5	35.7	148	2 S06019	isotocin 2 / neuro
14	5	35.7	149	2 T03501	conserved hypothet
15	5	35.7	149	2 A46132	c-Jun leucine zipp
16	5	35.7	154	2 A33896	isotocin 1 / neuro
17	5	35.7	154	2 T06396	soprenylated prot
18	5	35.7	161	2 S75584	hypothetical prote
19	5	35.7	161	2 S33491	hypothetical prote
20	5	35.7	188	2 C69388	probable dCTP deam
21	5	35.7	190	2 T05492	thaumatin homolog
22	5	35.7	191	2 T07540	probable light-rep
23	5	35.7	192	2 A72646	hypothetical prote
24	5	35.7	208	1 H70469	L-fucose-phospha
25	5	35.7	210	2 E84231	hypothetical prote
26	5	35.7	217	2 E96657	unknown protein, 8
27	5	35.7	219	2 A65076	ribose-5-phosphate
28	5	35.7	219	2 E85947	ribosephosphate is
29	5	35.7	244	2 T29399	hypothetical prote

30 5 35.7 252 1 A54677 homeotic protein g
31 5 35.7 255 2 F83575 conserved hypothet
32 5 35.7 256 1 A42768 homeotic protein g
33 5 35.7 271 2 D70915 hypothetical prote
34 5 35.7 272 1 UHHU2 interleukin-2 rece
35 5 35.7 275 2 G70022 iron(III) diclirat
36 5 35.7 281 2 B81747 ribosomal protein
37 5 35.7 282 2 G71484 probable S2 riboso
38 5 35.7 288 2 E70580 hypothetical prote
39 5 35.7 289 2 A83656 hypothetical prote
40 5 35.7 290 2 T34620 NADH dehydrogenase
41 5 35.7 290 2 T36205 hypothetical prote
42 5 35.7 305 2 A86046 sepQ [imported] -
43 5 35.7 308 2 D83452 probable cytochrom
44 5 35.7 309 2 T47685 probable RNA bindi
45 5 35.7 311 2 C84101 hypothetical prote

ALIGNMENTS

RESULT 1

E64437

probable dCTP deaminase (EC 3.5.4.13) MJ1102 [similarity] - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: E64437

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999

A:Accession: E64437

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-161 <BUL>

A:Cross-references: GB:U67553; GB:L77117; NID:gl591744; PIDN:AAB99105.1; PID:gl591747

C:Genetics:

A:Map position: FOR1043233-1043718

C:Superfamily: dCTP deaminase

C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 4e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SAVHDPGVEGRPEY 14

Db 110 SAVHDPGVEGRPEY 123

RESULT 2

A48350

DNA-binding protein - human herpesvirus 2

C:Species: human Herpesvirus 2

A>Note: host Homo sapiens (man)

C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 31-May-1996

C:Accession: A48350

R:Toh, Y.; Liu, Y.; Tanaka, S.; Mori, R.

Arch. Virol. 129, 183-196, 1993

A:Title: Nucleotide sequence of the major DNA-binding protein gene of herpes simplex

A:Reference number: A48350; MUID:93228441

A:Accession: A48350

A:Molecule type: DNA

A:Residues: 1-1197 <TOH>

A>Note: sequence extracted from NCBI backbone (NCBIN:129069, NCBIP:129070)

C:Genetics:

A:Map position: 0.375-0.405

C:Superfamily: herpesvirus DNA-binding protein

C:Keywords: DNA binding; zinc finger

F:499-512/region: zinc finger

Query Match 50.0%; Score 7; DB 1; Length 1197;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGYEGRP 12
 |||||
 Db 429 PGYEGRP 435

RESULT 3

T29427

Probable exonuclease - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 20-Jun-2000

C:Accession: T29427

R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1998

A:Reference number: Z20619

A:Accession: T29427

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-259 <PAR>

A:Cross-references: EMBL:AL031155; PIDN:CAA20072.1

C:Genetics:

A:Note: SC3A7.09

C:Superfamily: exodeoxyribonuclease III

Query Match 42.9%; Score 6; DB 2; Length 259;

Best Local Similarity 100.0%; Pred. No. 8.7;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DPGYEG 10

|||||

Db 83 DPGYEG 88

RESULT 4

E83263

hypothetical protein PA3054 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: E83263

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406: 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: E83263

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-634 <STO>

A:Cross-references: GB:AE004730; GB:AE004091; NID:99949154; PIDN:AG06442.1; GSPDB:GN001

C:Genetics:

A:Gene: PA3054

Query Match 42.9%; Score 6; DB 2; Length 634;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGYEGR 11

|||||

Db 504 PGYEGR 509

RESULT 5

S34839

collagen alpha 1(VI) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 13-Aug-1999

C:Accession: S34839; S31405; S32603; S31403; S32003

R:Bonaldo, P.; Piccolo, S.; Marvulli, D.; Volpin, D.; Marigo, V.; Bressan, G.M.

Matrix 13, 223-233, 1993

A:Title: Murine alpha-1(VI) collagen chain. Complete amino acid sequence and identifi

A:Reference number: S34839; MUID:93316904

A:Accession: S34839

A:Molecule type: mRNA

A:Residues: 1-1025 <BON>

A:Cross-references: EMBL:X66405; NID:950478; PIDN:CAA47032.1; PID:950479

R:Bonaldo, P.; Piccolo, S.; Marvulli, D.; Volpin, D.; Bressan, G.M.

submitted to the EMBL Data Library, May 1992

A:Reference number: S31403

A:Accession: S31405

A:Molecule type: DNA

A:Residues: 1-31 <BON>

A:Cross-references: EMBL:X66406; NID:951055; PIDN:CAA47033.1; PID:951056

R:Zhang, R.Z.; Pan, T.C.; Timpi, R.; Chu, M.L.

Biochem. J. 291, 787-792, 1993

A:Title: Cloning and sequence analysis of cDNAs encoding the alpha-1, alpha-2 and alp

A:Reference number: S32603; MUID:93256888

A:Accession: S32603

A:Molecule type: mRNA

A:Residues: 442-673, 'TL', 676-708, 'A', 710-942, 944-959, 'R', 961-1025 <ZHA>

A:Cross-references: EMBL:Z18271; NID:957955; PIDN:CAA79152.1; PID:957956

C:Genetics:

A:Gene: COL6A1

C:Superfamily: collagen alpha 1(VI) chain; von Willebrand factor type A repeat homolo

C:Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; heterotrim

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1025/Product: collagen alpha 1(VI) chain #status predicted <MAT>

F:34-213/Domain: von Willebrand factor type A repeat homology <VWAL>

Query Match 42.9%; Score 6; DB 2; Length 1025;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DPGYEG 10

|||||

Db 263 DPGYEG 268

RESULT 6

T04135

auxin-inducible protein homolog - maize (fragment)

C:Species: Zea mays (maize)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 18-Feb-2000

C:Accession: T04135

R:Gray, J.; Close, P.S.; Briggs, S.P.; Johal, G.S.

Cell 89, 25-31, 1997

A:Title: A novel suppressor of cell death in plants encoded by the L1s1 gene of maize

A:Reference number: Z15232; MUID:97248483

A:Accession: T04135

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-69 <GRA>

A:Cross-references: EMBL:U77346; NID:91935910; PIDN:AAC49677.1; PID:91935911

A:Experimental source: strain B73 inbred line

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 35.7%; Score 5; DB 2; Length 69;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GRPEY 14

|||||

Db 30 GRPEY 34


```

Query Match          35.7%; Score 5; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 AVHDP 6
        |||||
Db      99 AVHDP 103

RESULT 9
T03008
probable regulatory protein - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03008
R:Figueroa-Bossi, N.; Bossi, L.
submitted to the EMBL Data Library, June 1998
A:Reference number: Z14818
A:Accession: T03008
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-136 <FIG>
A:Cross-references: EMBL:AF001386; NID:g3294471; PIDN:AAC26077.1; PID:g3309514

Query Match          35.7%; Score 5; DB 2; Length 136;

```

	matches	conservative	mismatches	indels	gaps
Qy	10	GRPEY 14 	.		

```

DD      01 GRPEI 03

```

hypothetical protein sir0238 - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 A:Accession: S74547
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miya-
 o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S74547
 A:Status: preliminary
 A:Molecule type: D:NA
 A:Residues: 1-140 <RAN>
 A:Cross-references: EMBL:D80900; GB:AB001339; NID:q1651768; PIDN:BAA16699.1; PII:
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: *Synechocystis* hypothetical protein sir0238

Query Match	35.7%	Score 5;	DB 2;	Length 140;
Best Local Similarity	100.0%;	Pred. No. 71;		
Matches	5;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps	0;			

Qy 1 SAVHD 5
 |||||
 Db 61 SAVHD 65

RESULT 11
 T36136

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T36136

Received: 1999
Accession: 1999
R; Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream,
submitted to the EMBL Data Library, July 1999

A:Reference number: 221598

A:Accession: T36136
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-140 <SEE>
A:Cross-references: EMBL:AL096852; PIDN:CAB50994.1; GSPDB:GN00070; SCORDB:SCE19A.13c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCORDB:SCE19A.13c
C:Superfamily: Streptomyces coslicolor hypothetical protein SCE19A.13c

Query Match 35.7% Score 5; DB 2; Length 140;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 EGRPE 13
Db 34 EGRPE 38

RESULT 12

I51014
MHC class I protein - Nerodia sipedon (fragment)
C:Species: Nerodia sipedon
C>Date: 04-Sep-1997 #sequence_revision 07-Nov-1997 #text_change 23-Jul-1999
C:Accession: I51014
R:Grossberger, D.; Parham, P.
Immunogenetics 36, 166-174, 1992
A:Title: Reptilian class I major histocompatibility complex genes reveal conserved elements
A:Reference number: I50012; MUID:92307752
A:Accession: I51014
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-147 <GRO>
A:Cross-references: GB:M81099; NID:g213382; PIDN:AAA49389.1; PID:g213383
C:Genetics:
A:Gene: SC1
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology

Query Match 35.7% Score 5; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYEGR 11
Db 18 GYEGR 22

RESULT 13

S06019
Isotocin 2 / neurophysin precursor - white sucker
N:Contains: isotocin; neurophysin
C:Species: Catostomus commersoni (white sucker)
C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 16-Jul-1999
C:Accession: S06019
R:Figuerola, J.; Morley, S.D.; Heierhorst, J.; Krentler, C.; Lederis, K.; Richter, D.
EMBO J. 8, 2873-2877, 1989
A:Title: Two isotocin genes are present in the white sucker Catostomus commersoni both in the white sucker and in the yellow perch Perca flavescens
A:Reference number: S06018; MUID:90059870
A:Accession: S06019
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-148 <FTG>
A:Cross-references: GB:X15821; NID:g62634; PIDN:CAA34619.1; PID:g62635
C:Superfamily: oxytocin-neurophysin
C:Keywords: amidated carboxyl end; neuropeptide
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-29/product: isotocin #status predicted <ISO>
F:31-148/product: neurophysin #status predicted <NEU>
F:21-26/disulfide bonds: #status predicted
F:29/modified site: amidated carboxyl end (Gly) (amide in mature form from following gly

Query Match 35.7% Score 5; DB 2; Length 148;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYEGR 11
Db 94 GYEGR 98

RESULT 14

T03501
conserved hypothetical protein - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03501
R:Vicek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SBI
A:Reference number: Z14955; MUID:97404404
A:Accession: T03501
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-149 <VLC>
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16154.1; PID:g3128302
C:Genetics:
A:Map position: 1
C:Superfamily: hypothetical protein H11355

Query Match 35.7% Score 5; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YEGRP 12
Db 100 YEGRP 104

RESULT 15

A46132
C-Jun leucine zipper interactive (cDNA JZA-3) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 11-Apr-1995
C:Accession: A46132
R:Chevray, P.M.; Nathans, D.
Proc. Natl. Acad. Sci. U.S.A. 89, 5789-5793, 1992
A:Title: Protein interaction cloning in yeast: identification of mammalian proteins t
A:Reference number: A46132; MUID:92335183
A:Accession: A46132
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-149 <CHE>
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIP:108531)

Query Match 35.7% Score 5; DB 2; Length 149;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGYE 9
Db 74 DPGYE 78

Search completed: January 31, 2002, 13:20:14
Job time: 115 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:14 ; Search time 65.13 seconds
(without alignments)
4.837 Million cell updates/sec

Title: US-08-957-709-79

Perfect score: 14

Sequence: 1 AGVVDRDTGEVKV 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2.6/ptodata/2/laa/5A_COMB.pep.*
2: /cgn2.6/ptodata/2/laa/5B_COMB.pep.*
3: /cgn2.6/ptodata/2/laa/6A_COMB.pep.*
4: /cgn2.6/ptodata/2/laa/6B_COMB.pep.*
5: /cgn2.6/ptodata/2/laa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/2/laa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	147	2	US-08-824-405-6
2	6	42.9	333	2	US-08-896-083-2
3	6	42.9	333	2	US-08-896-083-4
4	6	42.9	333	3	US-09-110-910A-2
5	6	42.9	333	3	US-09-110-910A-4
6	5	35.7	32	2	US-08-023-980B-29
7	5	35.7	32	2	US-08-486-953A-24
8	5	35.7	40	1	US-08-314-596-3
9	5	35.7	40	1	US-08-314-596-4
10	5	35.7	40	1	US-08-314-596-9
11	5	35.7	40	1	US-08-314-596-17
12	5	35.7	40	1	US-08-314-596-23
13	5	35.7	40	1	US-08-314-596-24
14	5	35.7	40	1	US-08-320-982-3
15	5	35.7	40	1	US-08-320-982-4
16	5	35.7	40	1	US-08-320-982-9
17	5	35.7	40	1	US-08-320-982-17
18	5	35.7	40	1	US-08-320-982-23
19	5	35.7	40	1	US-08-320-982-24
20	5	35.7	40	3	US-08-819-037-3
21	5	35.7	40	3	US-08-819-037-4
22	5	35.7	40	3	US-08-819-037-9
23	5	35.7	40	3	US-08-819-037-17
24	5	35.7	40	3	US-08-819-037-23
25	5	35.7	40	3	US-08-819-037-24
26	5	35.7	81	2	US-08-469-412A-16
27	5	35.7	81	4	US-09-021-715-16

Sequence 8, Appli
Sequence 40, Appli
Sequence 39, Appli
Sequence 39, Appli
Sequence 39, Appli
Sequence 11, Appli
Sequence 27, Appli
Sequence 27, Appli
Sequence 27, Appli
Sequence 27, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 208, App
Sequence 189, App
Sequence 331, App
Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-824-405-6
; Sequence 6, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; TITLE OF INVENTION: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 147 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-824-405-6

Query Match 100.0%; Score 14; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FastSeq
CURRENT APPLICATION NUMBER: US/08/486,953A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/204,052
FILING DATE: 28-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/223002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/428-0200
TELEFAX: 617/428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-486-953A-24

Query Match 35.7%; Score 5; DB 2; Length 32;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGEVK 13
DB 9 TGEVK 13

RESULT 8
US-08-314-596-3
Sequence 3, Application US/08314596
Patent No. 5668292
GENERAL INFORMATION:
APPLICANT: SOMERVILLE, CHRIS
APPLICANT: VAN DE LOO, FRANK
TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,596
FILING DATE: 26-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 206905/1220
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-314-596-3

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
DB 19 VDRDY 23

RESULT 9
US-08-314-596-4
Sequence 4, Application US/08314596
Patent No. 5668292
GENERAL INFORMATION:

APPLICANT: SOMERVILLE, CHRIS
APPLICANT: VAN DE LOO, FRANK
TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,596
FILING DATE: 26-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 206905/1220
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-314-596-4

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
DB 19 VDRDY 23

RESULT 10
US-08-314-596-9
Sequence 9, Application US/08314596
Patent No. 5668292
GENERAL INFORMATION:

APPLICANT: SOMERVILLE, CHRIS
APPLICANT: VAN DE LOO, FRANK
TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,596
FILING DATE: 26-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 206905/1220
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-314-596-9

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
|||||
Db 19 VDRDY 23

RESULT 11
US-08-314-596-17
Sequence 17, Application US/08314596
Patent No. 5668292
GENERAL INFORMATION:
APPLICANT: SOMERVILLE, CHRIS
APPLICANT: VAN DE LOO, FRANK
TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,596
FILING DATE: 26-SEP-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 206905/1220
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-314-596-17

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
|||||
Db 19 VDRDY 23

RESULT 12
US-08-314-596-23
Sequence 23, Application US/08314596
Patent No. 5668292
GENERAL INFORMATION:
APPLICANT: SOMERVILLE, CHRIS
APPLICANT: VAN DE LOO, FRANK
TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,596
FILING DATE: 26-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 206905/1220
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-314-596-23

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
Db 19 VDRDY 23

RESULT 13
US-08-314-596-24
; Sequence 24, Application US/08314596
; Patent No. 5668292
; GENERAL INFORMATION:
; APPLICANT: SOMERVILLE, CHRIS
; APPLICANT: VAN DE LOO, FRANK
; TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
; TITLE OF INVENTION: PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/314,596
; FILING DATE: 26-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,596
; FILING DATE: 26-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 206905/1220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-314-596-24

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
Db 19 VDRDY 23

RESULT 14
US-08-320-982-3
; Sequence 3, Application US/08320982
; Patent No. 5801026
; GENERAL INFORMATION:
; APPLICANT: SOMERVILLE, CHRIS
; APPLICANT: VAN DE LOO, FRANK
; TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
; TITLE OF INVENTION: PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.

ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,982
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,596
; FILING DATE: 26-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 206905/1220
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-320-982-3

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
Db 19 VDRDY 23

RESULT 15
US-08-320-982-4
; Sequence 4, Application US/08320982
; Patent No. 5801026
; GENERAL INFORMATION:
; APPLICANT: SOMERVILLE, CHRIS
; APPLICANT: VAN DE LOO, FRANK
; TITLE OF INVENTION: USE OF PLANT FATTY ACYL HYDROXYLASES TO
; TITLE OF INVENTION: PRODUCE HYDROXYLATED FATTY ACIDS AND DERIVATIVES IN PLANTS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,982
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,596
; FILING DATE: 26-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 206905/1220
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-861-3000
 ; TELEFAX: 202-822-0944
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 40 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ;
 US-08-320-982-4

Query Match 35.7%; Score 5; DB 1; Length 40;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDRDY 8
 Db 19 VDRDY 23

Search completed: January 31, 2002, 13:15:15
 Job time: 96 sec

This Page Blank (uspto)


```

RESULT 2
072165 ID 072165 PRELIMINARY; PRT; 160 AA.
AC 072165
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DUTPASE HOMOLOG.
GN E3L.
OS orf virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Parapoxvirus.
OX NCBI_TaxID=10259;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D1701;
RA Ralga H.-J., Cottone R.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF056304; AAD03407.1; -.
DR HSSP: P06968; 1EUM.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD000946; dUTPase; 1.
SQ SEQUENCE 160 AA; 17175 MW; 4C9B13EFAEB0D0A3 CRC64;

Query Match 100.0%; Score 14; DB 12; Length 160;
Best Local Similarity 100.0%; Pred. No. 3.3e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNVGV 14
Db 81 AGVIDEDYRGNVGV 94

RESULT 3
093J44 ID 093J44 PRELIMINARY; PRT; 162 AA.
AC 093J44
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DUTPASE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kan L., Varanasi U., Zhu Y., Qi C., Reddy J.K.;
RT "Mouse dUTPase cloning and genomic structural analysis.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF091101; AAF74514.1; -.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD000946; dUTPase; 1.
SQ SEQUENCE 162 AA; 17414 MW; 88DF490EA9B890AD CRC64;

Query Match 100.0%; Score 14; DB 11; Length 162;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNVGV 14
Db 96 AGVIDEDYRGNVGV 109

RESULT 4
09CQ43 ID 09CQ43 PRELIMINARY; PRT; 162 AA.
AC 09CQ43
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 5031412106RIK PROTEIN.
GN 5031412106RIK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO, AND SPLEEN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lydons P., Marchionni L., Mashima J., Mazzarelli J., Momabaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wnshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK011407; BAB22591.1; -.
DR EMBL: AK003076; BAB22551.1; -.
DR MGD; MGI:1915007; 5031412106RIK.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD000946; dUTPase; 1.
SQ SEQUENCE 162 AA; 17384 MW; C9C3390EA9A4E0A4 CRC64;

Query Match 100.0%; Score 14; DB 11; Length 162;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNVGV 14
Db 96 AGVIDEDYRGNVGV 109

RESULT 5
09CU90 ID 09CU90 PRELIMINARY; PRT; 204 AA.
AC 09CU90
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 5133400F09RIK PROTEIN (FRAGMENT).
GN 5133400F09RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=INTESTINE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

```

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK017216; BAB30639.1; -;
 DR MGD: MGI:1918517; 5133400F09Rik.
 DR InterPro: IPR001428; dUTPase.
 DR Pfam: PF00692; dUTPase; 1.
 DR ProDom: PD000946; dUTPase; 1.
 FT NON_TER 1
 SQ SEQUENCE 204 AA; 21513 MW; A5907170BF71B37B CRC64;

Query Match 100.0%; Score 14; DB 11; Length 204;
 Best Local Similarity 100.0%; Pred. No. 4.2e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNGV 14
 Db 138 AGVIDEDYRGNGV 151
 |||||

RESULT 6
 P87630 PRELIMINARY; PRT; 147 AA.
 ID P87630
 AC P87630;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE 41KBP FRAGMENT FROM LEFT END OF GENOME.
 GN G2L.
 OS Cowpox virus (CPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10243;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRI-90;
 RX MEDLINE=97068532; PubMed=8963248;
 RA Saifonov P.F., Petrov N.A., Riabankina O.I., Totmenin A.V.,
 RA Shchelkunov S.N., Sandakchiev L.S.;
 RT "Genes of a circle of hosts for the cowpox virus.";
 RL Dokl. Akad. Nauk 349:829-833(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRI-90;
 RX MEDLINE=98229462; PubMed=9568042;
 RA Shchelkunov S.N., Saifonov P.F., Totmenin A.V., Petrov N.A.,
 RA Ryzankina O.I., Gutorov V.V., Kotwal G.J.;
 RT "The genomic sequence analysis of the left and right species-specific
 RT terminal region of a cowpox virus strain reveals unique sequences and
 RT a cluster of intact ORFs for immunomodulatory and host range
 RT proteins.";
 RL Virology 243:432-460(1998).
 DR EMBL: X94355; CAA64116.1; -;
 DR EMBL: Y11842; CAA72545.1; -;
 DR InterPro: IPR001428; dUTPase.
 DR Pfam: PF00692; dUTPase; 1.
 DR ProDom: PD000946; dUTPase; 1.
 SQ SEQUENCE 147 AA; 16255 MW; D562BC4F4827F1FA CRC64;

Query Match 71.4%; Score 10; DB 12; Length 147;

Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVIDEDYRGN 11
 Db 82 GVIDEDYRGN 91
 |||||

RESULT 7
 Q9JFF1 PRELIMINARY; PRT; 147 AA.
 ID Q9JFF1
 AC Q9JFF1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE TF2L.
 OS Vaccinia virus (strain Tian Tan).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OX NCBI_TaxID=10253;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIAN TAN;
 RA Jin Q., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,
 RA Jing D.Y., Yu W.H., Yuan J.S., Ma X.J.;
 RT "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF095889; AAF33892.1; -;
 DR InterPro: IPR001428; dUTPase.
 DR Pfam: PF00692; dUTPase; 1.
 DR ProDom: PD000946; dUTPase; 1.
 SQ SEQUENCE 147 AA; 16265 MW; DA961240B727FA2A CRC64;

Query Match 71.4%; Score 10; DB 12; Length 147;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVIDEDYRGN 11
 Db 82 GVIDEDYRGN 91
 |||||

RESULT 8
 Q9J5G5 PRELIMINARY; PRT; 145 AA.
 ID Q9J5G5
 AC Q9J5G5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ORF FPV038 DUTP PYROPHOSPHATASE VACCINIA F2L HOMOLOG.
 GN FPV038.
 OS Fowlpox virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Avipoxvirus.
 OX NCBI_TaxID=10261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20193820; PubMed=10729156;
 RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
 RT "The genome of fowlpox virus.";
 RL J. Virol. 74:3815-3831(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF198100; AAF44382.1; -;
 DR InterPro: IPR001428; dUTPase.
 DR Pfam: PF00692; dUTPase; 1.
 DR ProDom: PD000946; dUTPase; 1.
 SQ SEQUENCE 145 AA; 16252 MW; 94CF5F6963B47D08 CRC64;

Query Match 57.1%; Score 8; DB 12; Length 145;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRGNGV 14
 | | | | | | | |
 DB 82 DYRGNGV 89

RESULT 9
 Q90958 PRELIMINARY; PRT; 143 AA.
 AC Q90958;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BCDNA:LD08534 OR CG4584.
 GN S012L.
 DE G012L.
 OS Shope fibroma virus (strain Kasza) (SFV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Leorioipoxvirus
 OX NCBI_TaxID=10272;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KASZA;
 RX MEDLINE=84165064; PubMed=6323741;
 RA DeLange A.M., Macaulay C., Block W., Mueller T., McFadden G.;
 RT "Tumorigenic poxviruses: construction of the composite physical map of
 the Shope fibroma virus genome.";
 RL J. Virol. 50:408-416(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KASZA;
 RX MEDLINE=88322873; PubMed=2842947;
 RA Upton C., Macen J.L., Maranchuk R.A., DeLange A.M., McFadden G.;
 RT "Tumorigenic poxviruses: fine analysis of the recombination junctions
 in malignant rabbit fibroma virus, a recombinant between Shope fibroma
 virus and myxoma virus.";
 RL Virology 166:229-239(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KASZA;
 RX MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
 RA Proletti E.;
 RT "The complete DNA sequence of vaccinia virus.";
 RL Virology 179:247-266(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KASZA;
 RX MEDLINE=52074222; PubMed=1660196;
 RA Strayer D.S., Jerig H.H., O'Connor K.;
 RT "Sequence and analysis of a portion of the genomes of Shope fibroma
 virus and malignant rabbit fibroma virus that is important for viral
 replication in lymphocytes.";
 RL Virology 185:585-595(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-KASZA;
 RX MEDLINE=20032074; PubMed=10562495;
 RA Waller D.O., McFadden G., Evans D.H.;
 RT "The complete genome sequence of shope (Rabbit) fibroma virus.";
 RL Virology 264:319-343(1999).
 DR EMBL: AF170722; AAF17896.1; -;
 DR HSSP: P16088; IDUT.
 DR InterPro: IPR001428; dUTPase.
 DR Pfam: PF00692; dUTPase; 1.
 DR ProDom: PD000945; dUTPase; 1.
 SQ SEQUENCE 143 AA; 15425 MW; D6D3A92AA43AECDD1 CRC64;

Query Match 50.0%; Score 7; DB 12; Length 143;
 Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRGNGV 13
 | | | | | | | |
 DB 83 DYRGNGV 89

RESULT 10
 Q9V311 PRELIMINARY; PRT; 188 AA.
 AC Q9V311;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BCDNA:LD08534 PROTEIN.
 GN Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eurygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harrist N.L., Harvey D., Heiman T.J., Hernandez J.R., Heuvel J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mout R.S., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Scheeler F., Shen H.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhou Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu Q., Zheng L.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tsang G., Brokstein P., Frise E., Harvey D., Evans-Holm M.,
 RA Lewis S.E., Suh C., Rubin G.M.;
 RT "Full Length Drosophila melanogaster cDNA sequence.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003630; AAF53053.1; -;
 DR EMBL: AF160916; AAF46856.1; -;
 DR HSSP: P16088; IDUT.
 DR FlyBase: FBgn0027890; BCDNA:LD08534.


```
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR ProDom; PD000946; dUTPase; 1.
SQ SEQUENCE 188 AA; 19958 MW; C2C74EE89711B9F1 CRC64;

Query Match          50.0%; Score 7; DB 5; Length 188;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DEDYRGN 11
Db 98 DEDYRGN 104

RESULT 11
P95591
ID P95591 PRELIMINARY; PRT; 1060 AA.
AC P95591;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ROMPA (FRAGMENT).
GN OMPA.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDIAN TICK TYPHUS;
RA Raoult D., Fournier P.E., Roux V.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U83440; AAC35176.1; -.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
FT NON_TER 1
FT NON_TER 1060
SQ SEQUENCE 1060 AA; 110170 MW; 19E6FF23E555BBB6 CRC64;

Query Match          50.0%; Score 7; DB 2; Length 1060;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGVIDED 7
Db 671 AGVIDED 677

RESULT 12
P95592
ID P95592 PRELIMINARY; PRT; 1060 AA.
AC P95592;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ROMPA (FRAGMENT).
GN OMPA.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MJ;
RA Raoult D., Fournier P.E., Roux V.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U83443; AAC35179.1; -.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
FT NON_TER 1
FT NON_TER 1060

SQ SEQUENCE 1060 AA; 110162 MW; 2844797B5B7B112E CRC64;

Query Match          50.0%; Score 7; DB 2; Length 1060;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGVIDED 7
Db 671 AGVIDED 677

RESULT 13
P95593
ID P95593 PRELIMINARY; PRT; 1060 AA.
AC P95593;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ROMPA (FRAGMENT).
GN OMPA.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOROCCAN;
RA Raoult D., Fournier P.E., Roux V.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U83448; AAC35184.1; -.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
FT NON_TER 1
FT NON_TER 1060
SQ SEQUENCE 1060 AA; 110185 MW; 2E58DBC17B19B368 CRC64;

Query Match          50.0%; Score 7; DB 2; Length 1060;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGVIDED 7
Db 671 AGVIDED 677

RESULT 14
P95594
ID P95594 PRELIMINARY; PRT; 1060 AA.
AC P95594;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ROMPA (FRAGMENT).
GN OMPA.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEVEN MALISH;
RA Raoult D., Fournier P.E., Roux V.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U83453; AAC35189.1; -.
DR InterPro; IPR003858; rOmpA_rOmpB.
DR Pfam; PF02708; rOmpA_rOmpB; 1.
FT NON_TER 1
FT NON_TER 1060
SQ SEQUENCE 1060 AA; 110255 MW; 3DD48C98A231E76D CRC64;
```

Query Match 50.0%; Score 7; DB 2; Length 1060;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDED 7
Db 671 AGVIDED 677

RESULT 15

Q9YTTQ0 PRELIMINARY; PRT; 174 AA.
AC Q9YTTQ0;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ORF 16.
OS Ateeline herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_taxid=85618;
RN {}
RP SEQUENCE FROM N.A.
RC STRAIN=73;
RA Albrecht J.-C., Fleckenstein B.;
RT *Primary Structure of the Herpesvirus Ateles Genome.*;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083424; AAC95340.1; -;
DR InterPro; IPR00712; Bcl_2.
DR InterPro; IPR002475; BCL2_family.
DR Pfam; PF00452; Bcl-2; 1.
DR PROSITE; PS50062; BCL2_FAMILY; 1.
DR SMART; SM00337; BCL; 1.
SQ SEQUENCE 174 AA; 19492 MW; E4E476717CAB59DD CRC64;

Query Match 42.9%; Score 6; DB 12; Length 174;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDEDYR 9
Db 104 IDEDYR 109

Search completed: January 31, 2002, 13:37:52
Job time: 177 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:19 ; Search time 78.64 Seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-80

Perfect score: 14

Sequence: 1 AGVIDEDYRGNVGV 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	141	2	dutp pyrophosphata
2	14	100.0	164	2	dutp pyrophosphata
3	14	100.0	205	2	dutp pyrophosphata
4	10	71.4	144	1	retroviral protein
5	10	71.4	147	1	probable dutp pyro
6	10	71.4	147	2	C6L protein - vari
7	10	71.4	147	2	E2L protein - vari
8	10	71.4	147	2	hypothetical prote
9	7	50.0	188	3	nucleoside-triphos
10	7	50.0	1003	2	myosin IE heavy ch
11	6	42.9	174	2	hypothetical prote
12	6	42.9	178	2	probable dutp pyro
13	6	42.9	218	1	23K integral membr
14	6	42.9	239	2	probable PE protel
15	6	42.9	259	2	Trk potassium upda
16	6	42.9	271	2	dapB protein - Myc
17	6	42.9	289	2	glucose 1-dehydrog
18	6	42.9	304	2	probable purple ac
19	6	42.9	339	2	hypothetical prote
20	6	42.9	351	2	probable purple ac
21	6	42.9	394	2	threonine synthase
22	6	42.9	455	2	probable 2-isoprop
23	6	42.9	484	2	hypothetical prote
24	6	42.9	585	2	hypothetical prote
25	6	42.9	589	2	probable secreted
26	6	42.9	603	2	probable phosphogl
27	6	42.9	624	2	phosphoglycerate d
28	6	42.9	695	2	n-terminal acetyl
29	6	42.9	760	2	eye cell developme

30	6	42.9	819	2	T40527	hypothetical prote
31	6	42.9	849	2	A96592	hypothetical prote
32	6	42.9	878	2	B84977	alanine--trna liga
33	6	42.9	1268	2	T50252	probable transcrip
34	6	42.9	1391	2	S50608	hypothetical prote
35	6	42.9	1696	2	T24146	hypothetical prote
36	6	42.9	2869	2	T18518	apolipoprotein(a)
37	5	35.7	13	2	S14316	photosystem I 9K c
38	5	35.7	24	2	S21511	incFVI protein - E
39	5	35.7	74	2	T14687	hypothetical prote
40	5	35.7	76	2	B36850	Al9L protein - var
41	5	35.7	76	2	T28561	hypothetical prote
42	5	35.7	82	2	A72166	A20L protein - var
43	5	35.7	82	2	S15143	major coat protein
44	5	35.7	82	2	C83556	coat protein B of
45	5	35.7	103	2	T10397	protein gp16 - Org

ALIGNMENTS

RESULT 1
A46256
dutp pyrophosphatase (EC 3.6.1.23) - human
N:Alternate names: dutPase
C:Species: Homo sapiens (man)
C:Date: 08-Nov-1993 #sequence_revision 25-Aug-1995 #text_change 22-Jun-1999
C:Accession: A46256; A47548
R:McIntosh, E.M.; Ager, D.D.; Gadsden, M.H.; Haynes, R.H.
Proc. Natl. Acad. Sci. U.S.A. 89, 8020-8024, 1992
A:Title: Human dutp pyrophosphatase: cDNA sequence and potential biological importanc
A:Reference number: A46256; MUID:92390380
A:Accession: A46256
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-141 <MCI>
A:Cross-references: GB:M89913; NID:gi181843; PIDN:AAA58444.1; PID:gi181844
A:Note: sequence modified after extraction from NCBI backbone
R:McIntosh, E.M.; Ager, D.D.; Gadsden, M.H.; Haynes, R.H.
Proc. Natl. Acad. Sci. U.S.A. 90, 4328, 1993
A:Reference number: A38961
A:Contents: annotation; erratum
R:Strahler, J.R.; Zhu, X.; Hora, N.; Wang, Y.K.; Andrews, P.C.; Roseman, N.A.; Neel,
Proc. Natl. Acad. Sci. U.S.A. 90, 4991-4995, 1993
A:Title: Maturation stage and proliferation-dependent expression of dutPase in human
A:Reference number: A47548; MUID:93281681
A:Accession: A47548
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-141 <SPR>
A:Cross-references: GB:L11877; NID:g292876; PIDN:AAA36801.1; PID:g292877
C:Genetics:
A:Gene: GDB:DUT
A:Cross-references: GDB:135164; OMIM:601266
A:Map position: 15q15-15q21.1
C:Superfamily: retroviral proteinase
C:Keywords: hydrolase; phosphoprotein

Query Match 100.0%; Score 14; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNVGV 14
|||||
Db 75 AGVIDEDYRGNVGV 88

RESULT 2
G02777
dutp pyrophosphatase (EC 3.6.1.23) - human
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 22-Jun-1999

C:Accession: G02777
R:Cohen, D.; McIntosh, E.M.; Pearlman, R.E.
Submitted to the EMBL Data Library, July 1996
A:Reference number: H01715
A:Accession: G02777
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-164 <COH>
A:Cross-references: EMBL:U62891; NID:g1421817; PIDN:AAC51123.1; PID:g1421818
C:Genetics:
A:Gene: GDB:DUT
A:Cross-references: GDB:135164; OMIM:601266
A:Map position: 15q15-15q21.1
C:Superfamily: retroviral proteinase
C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 164;
Best Local Similarity 100.0%; Pred. No. 3.5e-08; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

QY 1 AGVIDEDYRGNVGV 14
|||||
Db 98 AGVIDEDYRGNVGV 111

RESULT 3
T10819
dUTP pyrophosphatase (EC 3.6.1.23) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
R:Chu, R.Y.; Lin, Y.L.; Rao, M.S.; Reddy, J.K.
Submitted to the EMBL Data Library, September 1998
A:Reference number: Z17173
A:Accession: T10819
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-205 <CHU>
A:Cross-references: EMBL:U64030; NID:g3550953; PID:g3550954
C:Superfamily: retroviral proteinase
C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 4.3e-08; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

QY 1 AGVIDEDYRGNVGV 14
|||||
Db 139 AGVIDEDYRGNVGV 152

RESULT 4
PRVZWR
retroviral proteinase-like protein - vaccinia virus (strain WR)
N:Alternate names: F2L protein
C:Species: vaccinia virus
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 18-Jun-1999
R:Slabaugh, M.B.; Roseman, N.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 4152-4155, 1989
A:Title: Retroviral protease-like gene in the vaccinia virus genome.
A:Reference number: A32907; MUID:89284577
A:Accession: A32907
A:Molecule type: DNA
A:Residues: 1-144 <SLA>
A:Cross-references: GB:M25392; NID:g335616; PIDN:AAA48238.1; PID:g335617
R:Roseman, N.A.; Slabaugh, M.B.
Virology 178, 410-418, 1990
A:Title: The vaccinia virus HindIII F fragment: nucleotide sequence of the left 6.2 kb.
A:Reference number: A36213; MUID:91020979
A:Accession: H36213

A:Molecule type: DNA
A:Residues: 1-144 <ROS>
A:Cross-references: EMBL:M34368
C:Superfamily: retroviral proteinase

Query Match 71.4%; Score 10; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.00062; Mismatches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVIDEDYRGN 11
|||||
Db 79 GVIDEDYRGN 88

RESULT 5
PRVZTF
probable dUTP pyrophosphatase (EC 3.6.1.23) - vaccinia virus (strain Copenhagen and A
N:Alternate names: dUTPase; F2L protein
C:Species: vaccinia virus
A:Variety: strain Ankara and Copenhagen
A:Note: host: Homo sapiens (man)
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 10-Dec-1999
C:Accession: G42506; T30780
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A:Title: Appendix to "The complete DNA sequence of vaccinia virus".
A:Reference number: A42501
A:Accession: G42506
A:Molecule type: DNA
A:Residues: 1-147 <GOE>
A:Cross-references: GB:M35027; NID:g335317; PIDN:AAA48015.1; PID:g335363
A:Experimental source: strain Copenhagen
R:Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A:Title: The complete DNA sequence of vaccinia virus.
A:Reference number: A42531; MUID:91021027
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
R:Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dörner, F.
Submitted to the EMBL Data Library, March 1997
A:Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) st
A:Reference number: Z20877
A:Accession: T30780
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-147 <ANT>
A:Cross-references: EMBL:U94848; PIDN:AAB96413.1
A:Experimental source: strain Ankara
C:Genetics:
A:Note: MVA030L
C:Superfamily: retroviral proteinase
C:Keywords: hydrolase; nucleotide metabolism; proteinase

Query Match 71.4%; Score 10; DB 1; Length 147;
Best Local Similarity 100.0%; Pred. No. 0.00063; Mismatches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVIDEDYRGN 11
|||||
Db 82 GVIDEDYRGN 91

RESULT 6
F36839
C6L protein - variola virus (strain India-1967)
C:Species: variola virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 23-Mar-2001
C:Accession: F36839
R:Blinov, V.M.
Submitted to GenBank, November 1992
A:Reference number: A36859

A;Accession: F36839
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-147 <BLI>
A;Cross-references: GB:X69198; NID:g456758; PIDN:CAA48967.1; PID:g297207
C;Superfamily: retroviral proteinase

Query Match 71.4%; Score 10; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 0.00063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVIDEDYRGN 11
|||||
DB 82 GVIDEDYRGN 91

RESULT 7
H72153
E2L protein - variola minor virus (strain Garcia-1966)
C;Species: variola minor virus
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
C;Accession: H72153
R;Shcheikunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopatin
submitted to GenBank, March 1998
A;Description: Analysis of the complete coding sequence of DNA of alastrim variola minor
A;Reference number: A72150
A;Accession: H72153
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-147 <SHC>
A;Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54626.1; PID:g5830587
A;Experimental source: strain Garcia-1966
C;Genetics:
A;Gene: E2L
C;Superfamily: retroviral proteinase

Query Match 71.4%; Score 10; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 0.00063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVIDEDYRGN 11
|||||
DB 82 GVIDEDYRGN 91

RESULT 8
T28464
hypothetical protein C6L - variola major virus
C;Species: variola major virus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T28464
R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Aubin
Nature 366, 748-751, 1993
A;Title: Potential virulence determinants in terminal regions of variola smallpox virus
A;Reference number: Z20488; MUID:94088747
A;Accession: T28464
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-147 <NAS>
A;Cross-references: EMBL:L22579; NID:g6233595; PIDN:AAA60774.1; PID:g438944
A;Experimental source: strain "Bangladesh-1975"
C;Superfamily: retroviral proteinase

Query Match 71.4%; Score 10; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 0.00063;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVIDEDYRGN 11
|||||
DB 82 GVIDEDYRGN 91

RESULT 9

JC7565
nucleoside-triphosphatase (EC 3.6.1.15) - fruit fly (*Drosophila melanogaster*)
N;Alternate names: deoxyridine-triphosphatase; dUTPase
C;Species: *Drosophila melanogaster*
C;Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C;Accession: JC7565

R;Fiser, A.; Vertessy, B.G.
Biochem. Biophys. Res. Commun. 279, 534-542, 2000
A;Title: Altered subunit communication in subfamilies of trimeric dUTPases.
A;Reference number: JC7565; MUID: 20568712
A;Accession: JC7565
A;Molecule type: mRNA
A;Residues: 1-188 <FIS>

C;Comment: This enzyme is essential in preventing uracil incorporation into DNA. It h
MP precursor for dUTP biosynthesis.
C;Keywords: hydrolase

Query Match 50.0%; Score 7; DB 3; Length 188;
Best Local Similarity 100.0%; Pred. No. 1.13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DEDYRGN 11
|||||
DB 98 DEDYRGN 104

RESULT 10

S33760
myosin IE heavy chain - slime mold (*Dictyostellium discoideum*)
C;Species: *Dictyostellium discoideum*
C;Date: 08-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 02-Feb-2001
C;Accession: S33760
R;Urrutia, R.A.; Jung, G.; Hammer III, J.A.
Biochim. Biophys. Acta 1173, 225-229, 1993

A;Title: The *Dictyostellium* myosin IE heavy chain gene encodes a truncated isoform tha
A;Reference number: S33760; MUID:93277957
A;Accession: S33760
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1003 <URR>
A;Cross-references: EMBL:L06805; NID:g167767; PIDN:AAA33201.1; PID:g167768
C;Genetics:
A;Introns: 77/3

C;Superfamily: brush border myosin heavy chain I; myosin motor domain homology
C;Keywords: nucleotide binding; p-loop
F;11-678/Domain: myosin motor domain homology <MMOT>
F;101-108/Region: nucleotide-binding motif A (p-loop)

Query Match 50.0%; Score 7; DB 2; Length 1003;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDED 7
|||||
DB 594 AGVIDED 600

RESULT 11

T42929
hypothetical protein - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T42929
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274

A:Accession: T42929
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-174 <ALB>
A:Cross-references: EMBL:AF083424; PIDN:AC95540.1
A:Experimental source: strain 73

Query Match 42.9%; Score 6; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDEYR 9
| | | | |
DB 104 IDEYR 109

RESULT 12
S26429
probable dUTP pyrophosphatase (EC 3.6.1.23) - fowl adenovirus 1
C:Species: Aviadonovirus galli (fowl adenovirus 1, CELO)
C:Date: 07-May-1993 #sequence_revision 02-Aug-1994 #text_change 12-Apr-1995
C:Accession: S26429
R:Akopian, T.A.; Kaverina, E.N.; Naroditsky, B.S.; Tikhonenko, T.I.
submitted to the EMBL Data Library, October 1992
A:Description: Avian adenovirus CELO: sequence of the DNA fragment (92-100%).
A:Reference number: S26428
A:Accession: S26429
A:Molecule type: DNA
A:Residues: 1-178 <AKO>
A:Cross-references: EMBL:Z17216
C:Superfamily: retroviral proteinase
C:Keywords: hydrolase

Query Match 42.9%; Score 6; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRCNV 12
| | | | |
DB 100 DYRCNV 105

RESULT 13
A43522
23K integral membrane protein - fluke (Schistosoma mansoni)
N:Alternate names: Sm23
C:Species: Schistosoma mansoni
C:Date: 28-Oct-1992 #sequence_revision 09-Aug-1996 #text_change 22-Jun-1999
C:Accession: A43522
R:Wright, M.D.; Henkle, K.J.; Mitchell, G.F.
J. Immunol. 144, 3195-3200, 1990
A:Title: An immunogenic M-r 23,000 integral membrane protein of Schistosoma mansoni worm
A:Reference number: A43522; MUID:90217533
A:Accession: A43522
A:Molecule type: mRNA
A:Residues: 1-218 <WRI>
A:Cross-references: GB:M34453; NID:g161029; PIDN:AAA29900.1; PID:g161030
A:Note: observations of molecular weight suggest this protein is at most lightly glycosylated
C:Superfamily: C9 antigen
C:Keywords: glycoprotein; transmembrane protein
F:1-13/Domain: intracellular #status predicted <CV1>
F:14-36/Domain: transmembrane #status predicted <EX1>
F:37-56/Domain: extracellular #status predicted <EX1>
F:57-77/Domain: transmembrane #status predicted <EX2>
F:78-82/Domain: intracellular #status predicted <EX2>
F:83-108/Domain: transmembrane #status predicted <EX2>
F:109-163/Domain: extracellular #status predicted <EX2>
F:164-208/Domain: transmembrane #status predicted <EX2>
F:209-218/Domain: intracellular #status predicted <EX3>
F:165/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 42.9%; Score 6; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRCNV 12
| | | | |
DB 152 DYRCNV 157

RESULT 14
C70684
probable PE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70684
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: C70684
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-239 <COL>
A:Cross-references: GB:Z81368; GB:AL123456; NID:g3261656; PIDN:CAB03719.1; PID:e27963
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PE

Query Match 42.9%; Score 6; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 RGNVG 14
| | | | |
DB 184 RGNVG 189

RESULT 15
D84343
TRK potassium uptake system protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84343
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: D84343
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <STO>
A:Cross-references: GB:AE004437; NID:g10581364; PIDN:AAG20112.1; GSPDB:GNO0138
C:Genetics:
A:Gene: trkA6

Query Match 42.9%; Score 6; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDEYR 9
| | | | |
DB 130 IDEYR 135

Search completed: January 31, 2002, 13:20:20
Job time: 121 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:15 ; Search time 65.13 seconds
(without alignments)
4.837 Million cell updates/sec

Title: US-08-957-709-80

Perfect score: 14

Sequence: 1 AGVIDEYRGNVG 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	147	2	US-08-824-405-12
2	14	100.0	164	2	US-08-824-405-2
3	14	100.0	252	2	US-08-824-405-4
4	10	71.4	144	2	US-08-824-405-13
5	6	42.9	271	5	PCT-US95-02455-2
6	6	42.9	760	1	US-08-195-152-2
7	5	35.7	43	4	US-08-905-223-286
8	5	35.7	78	2	US-08-612-788-18
9	5	35.7	78	2	US-08-763-528A-4
10	5	35.7	78	3	US-09-066-028-18
11	5	35.7	104	2	US-08-487-031-12
12	5	35.7	104	3	US-08-473-034-12
13	5	35.7	105	2	US-08-487-031-20
14	5	35.7	105	3	US-08-473-034-20
15	5	35.7	115	2	US-08-487-031-15
16	5	35.7	115	3	US-08-473-034-15
17	5	35.7	167	2	US-08-997-080-112
18	5	35.7	167	2	US-08-997-362-112
19	5	35.7	167	4	US-09-095-855-112
20	5	35.7	168	2	US-08-612-788-25
21	5	35.7	168	3	US-09-066-028-25
22	5	35.7	178	2	US-08-487-031-23
23	5	35.7	178	3	US-08-473-034-23
24	5	35.7	211	1	US-08-044-621B-25
25	5	35.7	211	1	US-08-709-912-8
26	5	35.7	211	2	US-09-047-370-8
27	5	35.7	229	4	US-09-248-528-11

Sequence 11, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 18, Appl
Sequence 6, Appl
Sequence 30, Appl
Sequence 30, Appl
Sequence 114, Appl
Sequence 114, Appl
Sequence 114, Appl
Sequence 114, Appl
Sequence 3, Appl
Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-824-405-12
: Sequence 12, Application US/08824405
: Patent No. 5962246
: GENERAL INFORMATION:
: APPLICANT: Ladner, Robert D.
: APPLICANT: Lynch, Frank
: APPLICANT: Caradonna, Salvatore J.
: TITLE OF INVENTION: Outphase, Its Isoforms, and
: TITLE OF INVENTION: Diagnostic and Other Uses
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert, Price & Rhoads
: STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2793
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/824,405
: FILING DATE: 26-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Bloom, Allen
: REGISTRATION NUMBER: 29,135
: REFERENCE/DOCKET NUMBER: 350163-101
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 609/520-3214
: TELEFAX: 609/520-3259
: TELEX:
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 147 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-824-405-12

Query Match 100.0%; Score 14; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-824-405-13

Query Match 71.4%; Score 10; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.00062;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GVIDEDYRGN 11
Db 79 GVIDEDYRGN 88

RESULT 5
PCT-US95-02455-2
; Sequence 2, Application PC/TUS9502455
; GENERAL INFORMATION:
; APPLICANT: Jacobs et al
; TITLE OF INVENTION: GENE FOR MYCOBACTERIAL
; DIAMINOPIMELIC ACID
; NUMBER OF INVENTION: 4
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amster, Rothstein & Ebenstein
; STREET: 90 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch 1.44 mb
; MEDIUM TYPE: storage diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02455
; FILING DATE: Not Yet Assigned
; CLASSIFICATION:
; PRIOR APPLICATION DATA: none
; APPLICATION NUMBER: none
; FILING DATE: none
; ATTORNEY/AGENT INFORMATION:
; NAME: George, Kenneth P.
; REGISTRATION NUMBER: 30,259
; REFERENCE/DOCKET NUMBER: 96700/358
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 271
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: dapB protein
; HYPOTHETICAL: no
```

```
; ANTI-SENSE:
; FRAGMENT TYPE: BCG
; ORIGINAL SOURCE: BCG
; ORGANISM: BCG
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION: none
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
; PCT-US95-02455-2

Query Match 42.9%; Score 6; DB 5; Length 271;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGVIDE 6
Db 19 AGVIDE 24

RESULT 6
US-08-195-152-2
; Sequence 2, Application US/08195152
; Patent No. 5679541
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,152
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
```

REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-195-152-2

Query Match 42.9%; Score 6; DB 1; Length 760;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YRGNVG 13
DB 616 YRGNVG 621

RESULT 7
US-08-905-223-286
; Sequence 286, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 286:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: -31..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 8.2
; OTHER INFORMATION: seq LECLLLYLAESSG/LR
US-08-905-223-286

Query Match 35.7%; Score 5; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GNVGV 14
DB 35 GNVGV 39

RESULT 8
US-08-612-788-18
; Sequence 18, Application US/08612788
; Patent No. 5837682
; GENERAL INFORMATION:
; APPLICANT: Folkman, M. Judah
; APPLICANT: O'Reilly, Micheal
; APPLICANT: Cao, Yihai
; APPLICANT: Sim, B. Kim Lee
; TITLE OF INVENTION: Angiostatin Fragments and Method of Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,788
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Watien, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0126
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: K3
US-08-612-788-18

Query Match 35.7%; Score 5; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YRGNV 12
DB 9 YRGNV 13

RESULT 9
US-08-763-528A-4
; Sequence 4, Application US/08763528A

Patent No. 5854221
GENERAL INFORMATION:
APPLICANT: Cao, Yihai
APPLICANT: Folkman, M. Judah
TITLE OF INVENTION: Endothelial Cell Proliferation Inhibitor
TITLE OF INVENTION: and Method of Use
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jones & Askew, LLP
STREET: 191 Peachtree Street, 37th Floor
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/763,528A
FILING DATE: 12-DEC-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Warren, William L.
REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 05940-0251
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3700
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
FEATURE:
NAME/KEY: Protein
LOCATION: 1..78
OTHER INFORMATION: /note= "Krangle 3 - Figure 3"
US-08-763-528A-4

Query Match 35.7%; Score 5; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YRGNV 12
Db 9 YRGNV 13

RESULT 10
US-09-066-028-18
Sequence 18, Application US/09066028
Patent No. 6024688
GENERAL INFORMATION:
APPLICANT: Folkman, M. Judah
APPLICANT: O'Reilly, Micheal
APPLICANT: Cao, Yihai
APPLICANT: Sim, B. Kim Lee
TITLE OF INVENTION: Angiostatin Fragments and Method of Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jones & Askew
STREET: 191 Peachtree Street, 37th Floor
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.

ZIP: 30303-1769
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/612,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren, William L.
REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 05213-0126
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: K3
US-09-066-028-18

Query Match 35.7%; Score 5; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YRGNV 12
Db 9 YRGNV 13

RESULT 11
US-08-487-031-12
Sequence 12, Application US/08487031
Patent No. 5977306
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Frank, Glenn R.
APPLICANT: Mika-Grieve, Marcia
APPLICANT: Tripp, Cynthia A.
TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P39
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,031
FILING DATE: 06-JUN-1995

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: 2618-34-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-031-12

Query Match 35.7%; Score 5; DB 2; Length 104;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DEDYR 9
DB 68 DEDYR 72

RESULT 12
US-08-473-034-12
Sequence 12, Application US/08473034
Patent No. 6114142
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Frank, Glenn R.
APPLICANT: Mika-Grieve, Marcia
APPLICANT: Tripp, Cynthia A.
TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P39
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/473.034
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: 2618-34
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-473-034-12

Query Match 35.7%; Score 5; DB 3; Length 104;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DEDYR 9
DB 68 DEDYR 72

RESULT 13
US-08-487-031-20
Sequence 20, Application US/08487031
Patent No. 5977306
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Frank, Glenn R.
APPLICANT: Mika-Grieve, Marcia
APPLICANT: Tripp, Cynthia A.
TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P39
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/487.031
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: 2618-34-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 105 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-031-20

Query Match 35.7%; Score 5; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DEDYR 9
DB 69 DEDYR 73

RESULT 14
US-08-473-034-20
Sequence 20, Application US/08473034
Patent No. 6114142
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Frank, Glenn R.
APPLICANT: Mika-Grieve, Marcia
APPLICANT: Tripp, Cynthia A.
TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P39
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh

```
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/473,034
; APPLICATION NUMBER: US/08/473,034
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: 2618-34
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 105 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-473-034-20

Query Match          35.7%; Score 5; DB 3; Length 105;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DEDYR 9
Db 69 DEDYR 73

RESULT 15
US-08-487-031-15
; Sequence 15, Application US/08487031
; Patent No. 5977306
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Frank, Glenn R.
; APPLICANT: Mika-Grieve, Marcia
; APPLICANT: Tripp, Cynthia A.
; TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P39
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,031
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: 2618-34-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
```

```
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-031-15
```

```
Query Match          35.7%; Score 5; DB 2; Length 115;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DEDYR 9
Db 79 DEDYR 83

Search completed: January 31, 2002, 13:15:16
Job time: 97 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:14 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-80

Perfect score: 14

Sequence: 1 AGVIDEYRGNGV 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_ll01.*

1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	14	100.0	14	AAW72856	Human dUTPase urid
2	14	100.0	141	AAW70144	Human dUTPase prot
3	14	100.0	184	AAW30280	Human dUTPase (nuc
4	14	100.0	252	AAW30281	Human dUTPase (mit
5	9	64.3	63	AAW58469	Lung cancer associ
6	6	42.9	219	AAW44032	Arabidopsis thalia
7	6	42.9	219	AAW50369	Arabidopsis thalia
8	6	42.9	221	AAW44031	Arabidopsis thalia
9	6	42.9	221	AAW50368	Arabidopsis thalia
10	6	42.9	227	AAW50367	Arabidopsis thalia
11	6	42.9	241	AAW44030	Arabidopsis thalia

12	6	42.9	271	16	AAW79946	ECG diaminopimelic
13	6	42.9	317	21	AAW50357	Arabidopsis thalia
14	6	42.9	326	21	AAW50356	Arabidopsis thalia
15	6	42.9	328	21	AAW50355	Arabidopsis thalia
16	6	42.9	345	22	AAW83037	S. epidermidis ope
17	6	42.9	432	21	AAW15617	Arabidopsis thalia
18	6	42.9	514	21	AAW28127	Arabidopsis thalia
19	6	42.9	514	21	AAW40020	Arabidopsis thalia
20	6	42.9	603	21	AAW28126	Arabidopsis thalia
21	6	42.9	603	21	AAW40019	Arabidopsis thalia
22	6	42.9	612	21	AAW28125	Arabidopsis thalia
23	6	42.9	612	21	AAW40018	Arabidopsis thalia
24	6	42.9	624	21	AAW15616	Arabidopsis thalia
25	6	42.9	636	21	AAW15615	Arabidopsis thalia
26	6	42.9	760	18	AAW29490	Arabidopsis thalia
27	5	35.7	25	22	AAW19319	Programmed cell de
28	5	35.7	25	22	AAW32103	Peptide #5753 enco
29	5	35.7	31	21	AAW12612	Peptide #6140 enco
30	5	35.7	39	22	AAW04675	Zea mays protein f
31	5	35.7	43	20	AAW13002	Haem protein relat
32	5	35.7	61	22	AAW21869	Human secreted pro
33	5	35.7	61	22	AAW38194	Peptide #8303 enco
34	5	35.7	63	21	AAW50020	Peptide #12231 enc
35	5	35.7	66	22	AAW35411	Arabidopsis thalia
36	5	35.7	78	17	AAW07555	Peptide #9448 enco
37	5	35.7	82	22	AAW20776	Human kringle 3.
38	5	35.7	82	22	AAW36136	Peptide #7210 enco
39	5	35.7	83	21	AAW01908	Human plasminogen
40	5	35.7	86	22	AAW15838	Peptide #2272 enco
41	5	35.7	86	22	AAW28349	Peptide #2386 enco
42	5	35.7	86	22	AAW03576	Peptide #2258 enco
43	5	35.7	97	21	AAW61744	Arabidopsis thalia
44	5	35.7	99	21	AAW09859	Arabidopsis thalia
45	5	35.7	99	21	AAW25224	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAW72856
ID AAW72856 standard; Peptide; 14 AA.
XX
AC AAW72856;
XX
DT 01-MAR-1999 (first entry)
XX
DE Human dUTPase uridine-binding motif.
XX
KW Polymerase enhancing factor; PEF; dUTPase; PCR; amplification;
KW sequencing; replication; human.
XX
OS Homo sapiens.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
XX
PR 21-MAR-1997; 97US-0822774.
XX
XX (STRA-) STRATAGENE.
XX
XX Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes -
XX PT improve the polymerisation activity of nucleic acid polymerases, for
XX PT use in amplification, sequencing and replication

PS Claim 71; Page 47; 161pp; English.

CC This is the uridine-binding motif of human dUTPase. Sequences

CC are provided (see AAW72849-57) of the uridine-binding motifs of the

CC dUTPases and dCTP deaminases of *Pyrococcus furiosus* (see also

CC AAW72847), *Methanococcus jannaschii*, *Desulfurolobus ambivalens*,

CC *Escherichia coli*, yeast, human and herpesvirus; a consensus (see

CC AAW72848) is also provided. A claimed method of enhancing a nucleic

CC acid polymerase reaction comprises performing the reaction in the

CC presence of one or more of the following: a polymerase enhancing

CC factor (PEF), a dUTPase, a protein that turns-over dUTP and a

CC protein having one or more of the sequences provided in AAW72848-57.

CC A claimed protein having PEF activity comprises one or more of

CC sequences given in AAW72848-57. Kits are provided for replicating

CC nucleic acids, for site-directed mutagenesis, for nucleic acid

CC sequencing or for amplification (preferably PCR or RT-PCR).

XX Sequence 14 AA;

XX

Query Match 100.0%; Score 14; DB 19; Length 14;

Best Local Similarity 100.0%; Pred. No. 7.3e-09; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 0;

QY 1 AGVIDEDYRGNGV 14

DB 1 agvidedyrgngv 14

RESULT 2

AAW70144

ID AAR70144 standard; Protein; 141 AA.

XX

AC AAR70144;

XX

DT 23-SEP-1995 (first entry)

DE Human dUTPase protomer.

DE

DE Deoxyuridine-triphosphate-nucleotidohydrolase; dUTPase;

KW antimicrobial; anticancer; cytostatic; screening.

KW

XX Homo sapiens.

OS

XX CA2126001-A.

PN

XX 28-JAN-1995.

PD

XX 16-JUN-1994; 94CA-2126001.

PF

XX 27-JUL-1993; 93US-0097220.

PR

XX (ALLX) ALLELIX BIOPHARMACEUTICALS INC.

PA

XX Climie S, Vandenberg E;

PI

XX WPI: 1995-115928/16.

DR

DR N-PSDB; AAQ83219.

XX

XX Pure native human deoxyuridine triphosphate nucleotidohydrolase

PT - used to identify inhibitors, potentially useful as

PT antimicrobial and anticancer agents

PT

XX Disclosure; Fig.1; 18pp; English.

PS

XX DNA encoding human dUTPase (given in AAQ83219) was cloned into pUC19

XX and recombinants used to transform *E. coli* BW 286. A transformant

CC selected for growth at 39 deg was cultured overnight. Recombinant

CC human dUTPase was purified and shown to be identical in sequence to

CC the natural protomeric peptide (AAW70144).

CC

XX Sequence 141 AA;

SQ

Query Match 100.0%; Score 14; DB 16; Length 141;

Best Local Similarity 100.0%; Pred. No. 6.1e-08; Indels 0; Gaps 0;

Matches 14; Conservative 0; Mismatches 0;

QY 1 AGVIDEDYRGNGV 14

DB 75 agvidedyrgngv 88

RESULT 3

AAW30280

ID AAW30280 standard; Protein; 164 AA.

XX

AC AAW30280;

XX

DT 14-APR-1998 (first entry)

DE Human dUTPase (nuclear form).

DE

DE dUTPase; DUT-N; human; cell proliferation; inhibitor; cytostatic;

KW antiviral; antibacterial; antifungal; antibiotic.

KW

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FH Modified-site 11

FT /note= "O-phosphorylated"

FT 1..5

FT Peptide

FT /note= "N-terminal peptide unique to DUT-N"

XX WO9736916-A1.

PN

XX 09-OCT-1997.

PD

XX 26-MAR-1997; 97WO-US04886.

PF

XX 26-MAR-1997; 97US-0824405.

PR

XX 29-MAR-1996; 96US-0014748.

PR

XX (UYNE-) UNIV NEW JERSEY.

PA

XX Caradonna SJ, Ladner RD, Lynch F;

PI

XX WPI: 1997-503040/46.

DR

DR N-PSDB; AAT90987.

XX

XX Nucleic acid encoding two isoform(s) of human dUTPase - used in

PT assays for assessing status of cell proliferation, effect of

PT anticancer agents and progress of treatment

PT

XX Claim 6; Fig 1A-B; 89pp; English.

PS

XX This polypeptide comprises the nuclear form, designated DUT-N, of

XX human deoxyuridine triphosphate nucleotidohydrolase (dUTPase). Its

CC amino acid sequence was deduced from a T-cell cDNA clone (see

CC AAT90987). A single gene codes for 2 isoforms of human dUTPase, the

CC nuclear form (DUT-N) and a cytoplasmic form (DUT-M, see AAW30281)

CC which is targeted to the mitochondria. The isoforms arise by the

CC use of alternative 5' exons which then join in a common sequence.

CC The DUT-N isoform is produced at low or undetectable levels during

CC non-proliferative periods of the cell cycle, and can be used as a

CC marker for cellular proliferation. It is phosphorylated by p34cdc2

CC at a Ser residue within the consensus sequence for cyclin-dependent

CC kinase phosphorylation. Measuring the amount of dUTPase in a cell

CC can be used to determine its proliferative status (or changes in

CC this status), particularly in nutrient-deprived or neoplastic cells.

CC (claimed). This method can be used for determining the effect of

CC anticancer agents (A) on cells, or monitoring the effect of (A)

CC (claimed). Agents that inhibit dUTPase are potential

CC antineoplastic, antiviral, antibacterial and antifungal agents.

CC The dUTPase can also be used to raise antibodies for assaying the

CC enzyme and for drug screening.

CC

XX SQ Sequence 164 AA;

Query Match 100.0%; Score 14; DB 18; Length 164;
Best Local Similarity 100.0%; Pred. No. 7e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNVGV 14
IIIIIIIIIIIIIIIIIIII

Db 98 agvidedyrgnvgv 111

RESULT 4
AAW30281
ID AAW30281 standard; Protein; 252 AA.
XX
AC AAW30281;
XX
DT 14-APR-1998 (first entry)
XX
DE Human dUTPase (mitochondrial form).
XX
KW dUTPase; DUT-M; human; cell proliferation; inhibitor; cytostatic;
KW antiviral; antibacterial; antifungal; antibiotic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 70..93
FT /note= "DUT-M unique sequence 5' to common Glu
FT residue of DUT-M and DUT-N"
XX
XX W09736916-A1.
XX
PD 09-OCT-1997.
XX
XX 26-MAR-1997; 97WO-US04886.
XX
XX 26-MAR-1997; 97US-0824405.
XX 29-MAR-1996; 96US-0014748.
XX
XX (UYNE-) UNIV NEW JERSEY.
XX
XX Caradonna SJ, Ladner RD, Lynch F;
XX
XX WPI; 1997-503040/46.
XX N-PSDB; AAT90989.
XX
XX Nucleic acid encoding two isoform(s) of human dUTPase - used in
XX assays for assessing status of cell proliferation, effect of
XX anticancer agents and progress of treatment
XX
XX Claim 6; Fig 2A-B; 89pp; English.

CC This polypeptide comprises the mitochondrial form, designated
CC dUT-M, of human deoxyuridine triphosphate nucleotidohydrolase
CC (dUTPase). Its amino acid sequence was deduced from a fibroblast
CC cDNA clone (see AAT90989). A single gene codes for 2 isoforms of
CC human dUTPase, the nuclear form (DUT-N, see AAW30280) and the
CC cytoplasmic form (DUT-M) which is targeted to the mitochondria.
CC The isoforms arise by the use of alternative 5' exons which then
CC join in a common sequence. The DUT-M isoform is detectable in
CC proliferating and non-proliferating cells, but is produced in
CC increased amounts during periods of non-proliferation in certain
CC tumours. Unlike DUT-N, it is not phosphorylated in vivo.
CC Measuring the amount of dUTPase in a cell can be used to determine
CC its proliferative status (or changes in this status), particularly
CC in nutrient-deprived or neoplastic cells (claimed). This method
CC can be used for determining the effect of anticancer agents (A) on
CC cells, or monitoring the effect of (A) (claimed). Agents that
CC inhibit dUTPase are potential antineoplastic, antiviral,
CC antibacterial and antifungal agents. The dUTPase can also be used

CC to raise antibodies for assaying the enzyme and for drug screening.

XX SQ Sequence 252 AA;

Query Match 100.0%; Score 14; DB 18; Length 252;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVIDEDYRGNVGV 14
IIIIIIIIIIIIIIIIIIII

Db 186 agvidedyrgnvgv 199

RESULT 5
AAB58469
ID AAB58469 standard; Protein; 63 AA.
XX
AC AAB58469;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polypeptide sequence SEQ ID 807.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardiotactic; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX
XX Homo sapiens.
XX OS
XX WO200055180-A2.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05918.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX
XX Ruben SM;
XX
XX WPI; 2000-587514/55.
XX N-PSDB; AAF18345.
XX
XX Lung cancer associated gene sequences, referred to as lung cancer
XX antigens, useful for treatment, prevention, and diagnosis of disorders
XX such as lung cancer -
XX
XX Claim 11; Page 1347; 1425pp; English.

CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer.
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.

```

SQ Sequence 63 AA:
Query Match 64.3%; Score 9; DB 21; Length 63;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 6 EDYRGNGV 14
Db 17 edyrngv 25
|||||
|||||

RESULT 6
AAG44032
ID AAG44032 standard; Protein; 219 AA.
XX
AC AAG44032;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55105.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134376.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
99US-0138094.
99US-0138540.
99US-0138847.
99US-0139119.
99US-0139452.
99US-0139453.
99US-0139492.
99US-0139454.
99US-0139455.
99US-0139456.
99US-0139457.
99US-0139458.
99US-0139459.
99US-0139460.
99US-0139461.
99US-0139462.
99US-0139463.
99US-0139750.
99US-0139763.
99US-0139817.
99US-0139899.
99US-0140353.
99US-0140354.
99US-0140695.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142055.
99US-0142390.
99US-0142803.
99US-0142920.
99US-0142977.
99US-0143542.
99US-0143624.
99US-0144005.
99US-0144085.
99US-0144086.
99US-0144325.
99US-0144331.
99US-0144332.
99US-0144333.
99US-0144334.
99US-0144335.
99US-0144352.
99US-0144352.
99US-0144804.
99US-0144814.
99US-0145086.
99US-0145088.
99US-0145088.
99US-0145087.
99US-0145089.
99US-0145192.
99US-0145145.
99US-0145218.
99US-0145224.
99US-0145276.
99US-0145913.
99US-0145918.
99US-0145919.
99US-0145951.
99US-0146386.
99US-0146388.
99US-0146389.
99US-0147038.
99US-0147204.
99US-0147302.
99US-0147192.
99US-0147260.
99US-0147303.
99US-0147416.

```

```
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.98; Score 6; DB 21; Length 219;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Oy 7 DYRGNY 12
|||||
Db 119 dyrgnv 124

RESULT 7
AAG50369
ID AAG50369 standard; Protein; 219 AA.
XX
AC AAG50369;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63822.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127482.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-0155139.
PR	08-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	12-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	13-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144095.	PR	07-OCT-1999;	99US-0158099.
PR	16-JUL-1999;	99US-0144096.	PR	08-OCT-1999;	99US-0158222.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	23-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160980.
PR	26-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	27-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145913.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145931.	PR	25-OCT-1999;	99US-0161405.
PR	28-JUL-1999;	99US-0145931.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161361.
PR	03-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147302.	PR	28-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			
PR	06-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			

Query Match 42.9%; Score 6; DB 21; Length 219;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRGNV 12
 Db 119 dyrgnv 124

RESULT 8
 AAG44031

ID AAG44031 standard; Protein; 221 AA.
XX AC AAG44031;
XX AC AAG44031;
DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 55104.
DE DE
XX Arabidopsis thaliana.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX PD
XX PF
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

```
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151130.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159233.
PR 13-OCT-1999; 99US-0159234.
PR 13-OCT-1999; 99US-0159235.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 221;
Best Local Similarity 100.0%; Pred No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRCNV 12
Db 121 dyrgnv 126

RESULT 9
AAG50368
IT AAG50368 standard; Protein; 221 AA.
AC XX
AC XX
DI 18-OCT-2000 (first entry)
XX XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63821.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX XX termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX XX
PD 06-SEP-2000.
XX PF
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 23-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
```



```
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9% Score 6; DB 21; Length 221;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 DYRGNV 12
Db 121 dyrgnv 126

RESULT 10
AAG50367
ID AAG50367 standard; Protein; 227 AA.
XX AC AAG50367;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 63820.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX
```

PN	EP1033405-A2.		
XX	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999;		99US-0121825.
PR	05-MAR-1999;		99US-0123180.
PR	09-MAR-1999;		99US-0123548.
PR	23-MAR-1999;		99US-0125788.
PR	23-MAR-1999;		99US-0126264.
PR	29-MAR-1999;		99US-0126785.
PR	01-APR-1999;		99US-0127462.
PR	06-APR-1999;		99US-0128234.
PR	16-APR-1999;		99US-0128714.
PR	19-APR-1999;		99US-0129845.
PR	21-APR-1999;		99US-0130077.
PR	23-APR-1999;		99US-0130449.
PR	23-APR-1999;		99US-0130510.
PR	28-APR-1999;		99US-0130891.
PR	30-APR-1999;		99US-0131449.
PR	04-MAY-1999;		99US-0132407.
PR	05-MAY-1999;		99US-0132484.
PR	06-MAY-1999;		99US-0132485.
PR	06-MAY-1999;		99US-0132486.
PR	07-MAY-1999;		99US-0132487.
PR	11-MAY-1999;		99US-0132863.
PR	14-MAY-1999;		99US-0134256.
PR	14-MAY-1999;		99US-0134218.
PR	14-MAY-1999;		99US-0134319.
PR	14-MAY-1999;		99US-0134321.
PR	18-MAY-1999;		99US-0134370.
PR	18-MAY-1999;		99US-0134768.
PR	20-MAY-1999;		99US-0134761.
PR	21-MAY-1999;		99US-0135154.
PR	21-MAY-1999;		99US-0135353.
PR	25-MAY-1999;		99US-0135629.
PR	25-MAY-1999;		99US-0136021.
PR	28-MAY-1999;		99US-0136392.
PR	01-JUN-1999;		99US-0136782.
PR	03-JUN-1999;		99US-0137222.
PR	04-JUN-1999;		99US-0137528.
PR	07-JUN-1999;		99US-0137502.
PR	08-JUN-1999;		99US-0137724.
PR	10-JUN-1999;		99US-0138094.
PR	10-JUN-1999;		99US-0138540.
PR	14-JUN-1999;		99US-0138847.
PR	16-JUN-1999;		99US-0139119.
PR	16-JUN-1999;		99US-0139452.
PR	17-JUN-1999;		99US-0139453.
PR	18-JUN-1999;		99US-0139492.
PR	18-JUN-1999;		99US-0139454.
PR	18-JUN-1999;		99US-0139455.
PR	18-JUN-1999;		99US-0139456.
PR	18-JUN-1999;		99US-0139457.
PR	18-JUN-1999;		99US-0139458.
PR	18-JUN-1999;		99US-0139459.
PR	18-JUN-1999;		99US-0139460.
PR	18-JUN-1999;		99US-0139461.
PR	18-JUN-1999;		99US-0139462.
PR	18-JUN-1999;		99US-0139463.
PR	18-JUN-1999;		99US-0139750.
PR	21-JUN-1999;		99US-0139763.
PR	21-JUN-1999;		99US-0139817.
PR	22-JUN-1999;		99US-0139899.
PR	23-JUN-1999;		99US-0140353.
PR	23-JUN-1999;		99US-0140354.
PR	24-JUN-1999;		99US-0140695.
PR	28-JUN-1999;		99US-0140823.
PR	29-JUN-1999;		99US-0140991.
PR	30-JUN-1999;		99US-0141287.
PR	01-JUL-1999;		99US-0141842.
PR	01-JUL-1999;		99US-0142154.
PR	02-JUL-1999;		99US-0142055.
PR	06-JUL-1999;		99US-0142390.
PR	08-JUL-1999;		99US-0142803.
PR	09-JUL-1999;		99US-0142920.
PR	12-JUL-1999;		99US-0142977.
PR	13-JUL-1999;		99US-0143542.
PR	14-JUL-1999;		99US-0143624.
PR	15-JUL-1999;		99US-0144005.
PR	16-JUL-1999;		99US-0144085.
PR	16-JUL-1999;		99US-0144086.
PR	19-JUL-1999;		99US-0144325.
PR	19-JUL-1999;		99US-0144331.
PR	19-JUL-1999;		99US-0144332.
PR	19-JUL-1999;		99US-0144333.
PR	19-JUL-1999;		99US-0144334.
PR	19-JUL-1999;		99US-0144335.
PR	20-JUL-1999;		99US-0144352.
PR	20-JUL-1999;		99US-0144632.
PR	20-JUL-1999;		99US-0144884.
PR	21-JUL-1999;		99US-0144814.
PR	21-JUL-1999;		99US-0145086.
PR	21-JUL-1999;		99US-0145088.
PR	22-JUL-1999;		99US-0145085.
PR	22-JUL-1999;		99US-0145087.
PR	22-JUL-1999;		99US-0145089.
PR	22-JUL-1999;		99US-0145192.
PR	23-JUL-1999;		99US-0145145.
PR	23-JUL-1999;		99US-0145218.
PR	23-JUL-1999;		99US-0145224.
PR	26-JUL-1999;		99US-0145276.
PR	27-JUL-1999;		99US-0145313.
PR	27-JUL-1999;		99US-0145318.
PR	28-JUL-1999;		99US-0145519.
PR	02-AUG-1999;		99US-0146386.
PR	02-AUG-1999;		99US-0146388.
PR	02-AUG-1999;		99US-0146389.
PR	03-AUG-1999;		99US-0147038.
PR	04-AUG-1999;		99US-0147204.
PR	05-AUG-1999;		99US-0147302.
PR	05-AUG-1999;		99US-0147192.
PR	06-AUG-1999;		99US-0147260.
PR	06-AUG-1999;		99US-0147303.
PR	09-AUG-1999;		99US-0147493.
PR	09-AUG-1999;		99US-0147935.
PR	10-AUG-1999;		99US-0148171.
PR	12-AUG-1999;		99US-0148319.
PR	13-AUG-1999;		99US-0148341.
PR	13-AUG-1999;		99US-0148565.
PR	16-AUG-1999;		99US-0148684.
PR	17-AUG-1999;		99US-0149368.
PR	18-AUG-1999;		99US-0149175.
PR	18-AUG-1999;		99US-0149426.
PR	20-AUG-1999;		99US-0149722.
PR	20-AUG-1999;		99US-0149723.
PR	23-AUG-1999;		99US-0149929.
PR	23-AUG-1999;		99US-0149902.
PR	25-AUG-1999;		99US-0149930.
PR	25-AUG-1999;		99US-0150566.
PR	26-AUG-1999;		99US-0150884.
PR	27-AUG-1999;		99US-0151065.
PR	27-AUG-1999;		99US-0151066.
PR	27-AUG-1999;		99US-0151080.
PR	30-AUG-1999;		99US-0151303.
PR	31-AUG-1999;		99US-0151438.
PR	01-SEP-1999;		99US-0151930.
PR	07-SEP-1999;		99US-0152363.
PR	10-SEP-1999;		99US-0152070.
PR	13-SEP-1999;		99US-0153758.
PR	15-SEP-1999;		99US-0154018.
PR	16-SEP-1999;		99US-0154039.

```
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 227;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRCNV 12
Db 127 dyrgnv 132

RESULT 11
AAG44030
ID AAG44030 standard; Protein; 241 AA.
XX AC
XX AAG44030;
XX AC
XX AC
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55103.
XX AC
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX AC
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX AC
XX AC
PD 06-SEP-2000.
XX AC
PF 25-FEB-2000; 2000EP-0301439.
XX AC
XX 25-FEB-1999; 99US-0121825.
```

PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145613.
 PR 27-JUL-1999; 99US-0145618.
 PR 27-JUL-1999; 99US-0145619.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 25-AUG-1999; 99US-0149930.
 PR 26-AUG-1999; 99US-0150566.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151430.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0152070.
 PR 13-SEP-1999; 99US-0153758.
 PR 16-SEP-1999; 99US-0154018.
 PR 20-SEP-1999; 99US-0154039.
 PR 22-SEP-1999; 99US-0154779.
 PR 23-SEP-1999; 99US-0155139.
 PR 24-SEP-1999; 99US-0155486.
 PR 28-SEP-1999; 99US-0155659.
 PR 29-SEP-1999; 99US-0156458.
 PR 04-OCT-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 42.3%; Score 6; DB 21; Length 241;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYGRNV 12
 |||||
 Db 141 dygrnv 146

RESULT 12

AAR79946

ID AAR79946 standard; Protein; 271 AA.

XX AAR79946;

XX 26-FEB-1996 (first entry)

DT BCG diaminopimelic acid synthetic enzyme DapB.

DE Dihydrodipicolinate-reductase; diaminopimelate-dehydrogenase; dapB;

XX diaminopimelic acid; peptidoglycan; mycobacteria; vaccine;

KW Mycobacterium tuberculosis; Mycobacterium avium;

KW Mycobacterium fortuitum; Mycobacterium leprae; Mycobacterium goodii;

KW Mycobacterium haemophilum; Mycobacterium paratuberculosis; BCG.

XX Mycobacterium bovis.

XX WO9523226-A1.

PN 31-AUG-1995.

XX 27-FEB-1995; 95WO-US02455.

XX 28-FEB-1994; 94US-0203190.

XX (YESH) UNIV YESHIVA EINSTEIN COLLEGE.

XX Cirillo JD, Jacobs WR;

XX

```

DR WPI; 1995-311535/40.
DX N-PSDB; AAT04154.
XX
PT Gene involved in the synthesis of di:amino:pinelic acid in
PT mycobacteria - and methods for inhibiting its expression to treat
PT mycobacterial infection
XX
PS Disclosure; Page 29-29; 48pp; English.
XX
CC A BGC dapB gene (see AAT04154) was isolated that encoded a
CC bifunctional enzyme (AAR79946) capable of catalyzing the
CC dihydrodipicolinate-reductase and diaminopimelate-dehydrogenase
CC reactions involved in diaminopimelic acid (DAP) biosynthesis.
CC Prodn. of mycobacterial DapB will allow the screening of
CC DapB-inhibitory (antimycobacterial) cpds.
XX
SQ Sequence 271 AA;

Query Match 42.9%; Score 6; DB 16; Length 271;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGVIDE 6
Db 19 agvide 24

RESULT 13
AAG50357
ID AAG50357 standard; Protein; 317 AA.
XX
AC AAG50357;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63806.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

```

```
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159233.
PR 13-OCT-1999; 99US-0159234.
PR 13-OCT-1999; 99US-0159235.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 317;
Best Local Similarity 100.0%; Pred No 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYRCNV 12
Db 110 dyrgnv 115

RESULT 14
AAG50356
ID AAG50356 standard; Protein: 326 AA.
XX AAG50356;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 63805.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63805.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 21-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126268.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 13-APR-1999; 99US-0129645.
PR 16-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
```

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 08-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 42.9%; Score 6; DB 21; Length 326;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 7 DYRGNV 12
| | | | |
Db 119 dyrgnv 124
RESULT 15
AAG50355
ID AAG50355 standard; Protein; 328 AA.
XX AC AAG50355;
XX DT
XX 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 63804.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134178.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134421.
PR 14-MAY-1999; 99US-0134570.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 328;
Best Local Similarity 100.0%; Pred. No. 46;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 DYRCNV 12
Db 121 dyrgnv 126
Search completed: January 31, 2002, 13:18:15
Job time: 176 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:10 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-76

Perfect score: 14

Sequence: 1 PRIVADGEGQTLI 14

Scoring table:

OLIGO 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_1101.*

1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	AAW72852	Desulfurolobus amb
2	14	100.0	173	AAE51078	Polypeptide encode
3	7	50.0	488	AAE20302	Arabidopsis appto
4	6	42.9	121	AAE40244	Human ORFX ORF8 po
5	6	42.9	159	AAE05583	Partial 36 kD anti
6	6	42.9	174	AAE81528	S. epidermidis ope
7	6	42.9	327	AAE05529	Soybean vestitone
8	6	42.9	327	AAE97832	Soybean vestitone
9	6	42.9	543	AAE20313	Soybean apptosis
10	6	42.9	585	AAE70143	BARL gene product.
11	6	42.9	587	AAE13383	Barrier protein.

12	6	42.9	587	13	AAE20109	BARL barrier prote
13	6	42.9	1001	21	AAE58677	Arabidopsis ranl c
14	6	42.9	1001	21	AAE58717	Arabidopsis ranl-1
15	6	42.9	1001	21	AAE58718	Arabidopsis ranl-2
16	5	35.7	7	17	AAW00484	Peptide epitope of
17	5	35.7	29	21	AAE23779	Entry vector pENTR
18	5	35.7	32	22	AAU01727	Human secreted pro
19	5	35.7	35	19	AAW79020	Rieske motif in DN
20	5	35.7	43	18	AAW14929	Homo sapiens liver
21	5	35.7	43	20	AAE22640	Liver neoplastic d
22	5	35.7	48	16	AAE85672	12 kD M.tuberculos
23	5	35.7	48	17	AAE92889	Mycobacterium 12 k
24	5	35.7	48	18	AAW18183	N-terminal sequenc
25	5	35.7	48	19	AAW75570	M. tuberculosis 12
26	5	35.7	50	20	AAE22641	Fragment of an alt
27	5	35.7	54	20	AAE22642	Fragment of an alt
28	5	35.7	54	22	AAW00024	GDP-D-mannose dehy
29	5	35.7	61	10	AAE92159	Central portion of
30	5	35.7	62	19	AAE69654	Galpha subunit swi
31	5	35.7	62	21	AAE20770	G alpha subunit sw
32	5	35.7	62	22	AAW17159	Peptide #3593 enco
33	5	35.7	62	22	AAE29650	Peptide #3687 enco
34	5	35.7	62	22	AAW04853	Peptide #3535 enco
35	5	35.7	91	20	AAE95318	Chlamydia major o
36	5	35.7	92	20	AAW95309	Chlamydia major o
37	5	35.7	94	20	AAW95308	Chlamydia major o
38	5	35.7	94	20	AAW95310	Chlamydia major o
39	5	35.7	94	20	AAW95311	Chlamydia major o
40	5	35.7	94	20	AAW95312	Chlamydia major o
41	5	35.7	94	20	AAW95315	Chlamydia major o
42	5	35.7	94	20	AAW95316	Chlamydia major o
43	5	35.7	94	22	AAW00025	GDP-D-mannose dehy
44	5	35.7	95	20	AAW95313	Chlamydia major o
45	5	35.7	95	20	AAW95314	Chlamydia major o

ALIGNMENTS

RESULT 1
AAW72852
ID AAW72852 standard; Peptide; 14 AA.
AC AAW72852;
DT 01-MAR-1999 (first entry)
XX Desulfurolobus ambivalens dCTP deaminase uridine-binding motif.
DE Desulfurolobus ambivalens dCTP deaminase uridine-binding motif.
XX Polymerase enhancing factor; PEF; dCTP deaminase; PCR;
KW amplification; sequencing; replication.
XX Desulfurolobus ambivalens.
OS
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
XX
PR 21-MAR-1997; 97US-0822774.
XX
(STRA-) STRATAGENE.
XX
PI Hansen CJ, Hogrefe H;
XX
DR WPI; 1998-542284/46.
XX
PT Polymerase enhancing factor proteins, extracts and complexes -
PT Improve the polymerisation activity of nucleic acid polymerases, for
XX use in amplification, sequencing and replication
XX

PS Claim 71; Page 47; 161pp; English.

XX This is the uridine-binding motif of the dCTP deaminase of
CC Desulfohalobium ambivalens. Sequences are provided (see AA72849-57)
CC of the uridine-binding motifs of dUTPases and dCTP deaminases of
CC Pyrococcus furiosus (see AA72847), Methanococcus jannaschii,
CC Desulfohalobium ambivalens, Escherichia coli, yeast, human and
CC herpesvirus; a consensus (see AA72848) is also provided. A claimed
CC method of enhancing a nucleic acid polymerase reaction comprises
CC performing the reaction in the presence of one or more of the
CC following: a polymerase enhancing factor (PEF), a dUTPase, a
CC protein that turns-over dUTP and a protein having one or more of
CC the sequences provided in AA72848-57. A claimed protein having PEF
CC activity comprises one or more of sequences given in AA72848-57.
CC Kits are provided for replicating nucleic acids, for site-directed
CC mutagenesis, for nucleic acid sequencing or for amplification
CC (preferably PCR or RT-PCR).

XX Sequence 14 AA;

Query Match 100.0%; Score 14; DB 19; Length 14;
Best Local Similarity 100.0%; Pred. NO. 3.9e-06; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

Qy 1 PTIVDAGFEGQITI 14
Db 1 PTIVDAGFEGQITI 14

RESULT 2

AA751078
ID AA751078 standard; Protein: 173 AA.

XX AA751078;

XX 27-MAY-1994 (first entry)

XX Polypeptide encoded by pDam-L3 plasmid fragment.

XX Ligase, thermostable; thermostability; Desulfohalobium ambivalens;
XX archaeobacteria; detection; ligation; ATP; adenosine triphosphate;
XX DE4217134-A.

XX 25-NOV-1993.

XX 23-MAY-1992; 92DE-4217134.

XX 23-MAY-1992; 92DE-4217134.

XX (BOEF) BOEHRINGER MANNHEIM GMBH
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Jarsch M, Kaletta C, Kessler C, Kletzin A, Rueger R;

XX WPI: 1993-378402/48.

XX N-PSDB; AAQ62300.

XX Thermostable ligase from archaeobacteria - and DNA coding for it,
XX useful for nucleic acid detection

XX Example 3; Figure 1; 26pp; German.

XX A sequence encoding a thermostable DNA ligase was cloned into the
XX plasmid pDam-L3. The DNA ligase is ATP dependent and is useful for
XX detecting nucleic acids by hybridising two oligonucleotides with
XX adjacent sequences of the target nucleic acid, ligating the two
XX oligonucleotides (using the ligase) and detecting the ligation
XX product. Analysis of the fragment revealed three other open reading
XX frames (ORF's). This is a polypeptide encoded by ORF3

XX Sequence 173 AA;

Query Match 100.0%; Score 14; DB 14; Length 173;
Best Local Similarity 100.0%; Pred. NO. 2.8e-07; Mismatches 0; Indels 0; Gaps 0;
Matches 14; Conservative 0;

Qy 1 PTIVDAGFEGQITI 14
Db 109 ptivdagfegqitl 122

RESULT 3

AA720302
ID AA720302 standard; Protein: 488 AA.

XX AA720302;

XX 29-MAY-2001 (first entry)

XX Arabidopsis apoptosis inducing factor-like protein.

XX Apoptosis inducing factor; AIF; monodehydroascorbate reductase;
XX male sterility; stress response; tissue culture; plant.

XX Arabidopsis thaliana.

XX WO200119994-A2.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000WO-US24859.

XX 13-SEP-1999; 99US-0153737.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Butler KH, Cahoon RE, Klein TM, Mazithulela G;

XX WPI: 2001-244804/25.

XX N-PSDB; AAF30431.

XX New polynucleotide encoding plant apoptosis inducing factor for
XX engineering male sterility, altering plant architecture, manipulating
XX stress response, and influencing growth of cells and tissues in culture
XX -
XX Claim 1(f); Page 48-50; 91pp; English.

XX The present sequence is that of a monodehydroascorbate reductase
XX like apoptosis inducing factor (AIF) of Arabidopsis thaliana
XX Landsberg estate. The sequence is predicted from the full-length
XX cDNA (see AAF30431). Of a clone isolated from a fertilised carpel
XX library, it shows 51.7% amino acid identity to previously
XX known monodehydroascorbate-reductase-like AIFs of Oryza sativa and
XX Brassica juncea. The invention relates to isolated nucleic acids
XX encoding plant AIFs. These are used in sense or antisense
XX constructs to alter levels of AIFs in plant cells, and hence to
XX alter apoptosis and eventually to control cell tissue culture
XX growth, facilitate studies of programmed cell death in plants,
XX increase the efficiency of gene transfer, help provide more stable
XX transformations, engineer male sterility, alter plant architecture
XX or manipulate stress response.

XX Sequence 488 AA;

Query Match 50.0%; Score 7; DB 22; Length 488;
Best Local Similarity 100.0%; Pred. NO. 9.1; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 8 FEGQITI 14
Db 267 fegqitl 273

RESULT 4
 AAB40244
 ID AAB40244 standard; Protein; 121 AA.
 XX
 AC AAB40244;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF8 polypeptide sequence SEQ ID NO:16.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN W0200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US08621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Leach M;
 PI
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC74453.
 DR
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 XX Claim 11; Page 431-432; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance

CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 121 AA;

Query Match 42.9%; Score 6; DB 21; Length 121;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDAGFE 9
 |||||
 Db 114 vdagfe 119

RESULT 5
 AAR05583
 ID AAR05583 standard; protein; 159 AA.
 XX
 AC AAR05583;
 XX
 DT 07-AUG-1990 (first entry)
 XX
 DE Partial 36 kD antigen of Mycoplasma hyopneumonia.
 XX
 KW Recombinant Mycoplasma hyopneumoniae antigens; mycoplasma
 KW pneumonia; vaccine; antibody; ds.
 XX
 OS Synthetic.
 XX
 PN EP359919-A.
 XX
 PD 28-MAR-1990.
 XX
 XX 28-JUN-1989; 89EP-0111748.
 PF
 XX 21-APR-1989; 89US-0341968.
 PR
 XX (MLTE) ML TECHN. VENTURES L.
 PA
 XX Faulds DH, Brooks E, Andrews WH, Lory C;
 PI
 XX WPI; 1990-092610/13.
 DR N-PSDB; AAQ03631.
 DR
 XX Recombinant Mycoplasma hyopneumoniae antigens - used in vaccines
 PT against mycoplasma pneumonia and in assays for detection or
 PT determination of antibody.
 XX
 PS Disclosure: Fig 32; 54pp; English.
 XX
 CC An expression vehicle containing the antigen coding sequence is
 CC capable of eliciting an antibody which recognises an epitope of
 CC an M. hyo antigen. The recombinant proteins can be used in a vaccine
 CC for protection against Mycoplasma pneumonia, partic in swine, or in
 CC an assay for detection of antibodies to M.hyo.
 CC See also AAQ03629-Q03634, AAQ04959-Q04963 and AAQ02474.

SQ Sequence 159 AA;

Query Match 42.9%; Score 6; DB 11; Length 159;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DAGFEG 10
 |||||
 Db 35 dagfeg 40

RESULT 6
 AAG81528
 ID AAG81528 standard; Protein; 174 AA.
 XX

AC AAG81528;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:150.
 XX
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 XX vaccination; endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 FN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 200WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 XX WPI: 2001-316495/33.
 XX
 DR N-PSDB; AAH52378.
 XX
 DT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 XX useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 18; Page 84; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120 from Staphylococcus epidermidis
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of the
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH53091 to
 CC AAH53098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 174 AA;
 XX
 Query Match 42.9%; Score 6; DB 22; Length 174;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 4 VDAGFE 9
 Db 145 vdagfe 150
 |||||
 RESULT 7
 AAAY05529
 ID AAAY05529 standard; Protein: 327 AA.
 XX
 AC AAAY05529;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 KW Soybean vestitone reductase.
 DE

XX Vestitone reductase; soybean; isoflavone; transgenic plant;
 KW antihemolytic; hypocholesterolemic; hypolipidemic; oestrogenic;
 KW antitumour; antifungal.
 XX
 OS Glycine max.
 XX
 FN WO9914351-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 17-SEP-1998; 98WO-US19553.
 XX
 PR 17-SEP-1997; 97US-0931668.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Fader GM;
 XX
 XX WPI: 1999-244040/20.
 XX
 DR N-PSDB; AAX25131.
 XX
 PT Isolated nucleic acid encoding enzymes for isoflavone biosynthesis
 XX
 PS Claim 9; Page 33-39; 42pp; English.
 XX
 CC The present sequence is soybean vestitone reductase, encoded by
 CC the cDNA insert (see AAX25131) of an EST clone, ser.pk0016.c8,
 CC isolated from a soybean root cDNA library following a BLASTX
 CC search. The sequence shows similarity to Medicago sativa vestitone
 CC reductase. The invention relates to isolated nucleic acids
 CC encoding soybean enzymes that catalyse steps in the biosynthesis of
 CC isoflavones i.e. chalcone isomerase (see AAY05527). Isoflavone
 CC reductase (see AAY05528) and vestitone reductase encoding all or
 CC substantial portion of the enzymes, in sense or antisense orientation,
 CC which express one of the enzymes, in sense or antisense orientation,
 CC levels of the enzyme in a transformed cell. The method is all
 CC used to alter isoflavone levels in plants and seeds, particularly
 CC of food plants. High accumulation of isoflavones may have flavour
 CC or health benefits. Isoflavones, generally as conjugates with
 CC sugars or acids, are known to have antihemolytic, antifungal,
 CC oestrogenic, tumor-suppressing, antihypercholesterolemic and
 CC hypolipidemic activities. Since the specified enzymes are involved
 CC late in isoflavone synthesis, they are unlikely to interfere with
 CC other parts of the phenylpropanoid pathway (leading to lignin,
 CC anthocyanins and flavones).
 XX
 SQ Sequence 327 AA;
 XX
 Query Match 42.9%; Score 6; DB 20; Length 327;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 4 VDAGFE 9
 Db 300 vdagfe 305
 |||||
 RESULT 8
 AAAY97832
 ID AAAY97832 standard; Protein: 327 AA.
 XX
 AC AAAY97832;
 XX
 DT 29-AUG-2000 (first entry)
 XX
 DE Soybean vestitone reductase, encoded by cDNA clone sre.pk0016.c8.
 KW Soybean; vestitone reductase; root elongation zone; EST;
 KW expressed sequence tag; isoflavone biosynthesis;
 KW phenylpropanoid biosynthetic pathway; transgenic plant; antihemolytic;

KW antifungal; oestrogenic; tumour suppressor; hypolipidaemic;
KW hypocholesterolaemic.

OS Glycine max.

XX US6054636-A.

XX 25-APR-2000.

XX 17-SEP-1998; 98US-0154874.

XX 17-SEP-1997; 97US-0931668.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Fader GM;

XX WPI; 2000-338516/29.

XX N-PSDB; AAA38426.

XX New polynucleotide encoding chalcone isomerase for altering the levels
PT of plant isoflavone biosynthetic enzyme in a host cell and for
PT producing transgenic plants and antibodies

XX Example 2; Fig 4; 25pp; English.

XX This sequence represents soybean vestitone reductase, encoded by cDNA
CC clone sre.pK0016.c8. This cDNA was identified in a soybean root
CC elongation zone cDNA library, and is the first expressed sequence
CC tag (EST) identified for soybean chalcone isomerase. Vestitone
CC reductase, chalcone isomerase (AAY97630) and isoflavone reductase
CC (AAY97831) are component enzymes of the phenylpropanoid biosynthetic
CC pathway which is responsible for the biosynthesis of isoflavones from
CC phenylalanine in leguminous plants. The isoflavones found in soybeans are
CC thought to provide several health benefits to humans. Soybean isoflavones
CC have antithaemolytic, antifungal, oestrogenic, tumour suppressing,
CC hypolipidaemic and serum cholesterol-lowering effects. However, certain
CC isoflavones, such as isoflavone, flavone, anthocyanins and flavanols
CC adversely affect the flavour of certain soy food products. The nucleic
CC acids of the invention (AAA38424-A38426) may be used to generate
CC transgenic plants in which expression of chalcone isomerase, isoflavone
CC reductase and/or vestitone reductase is altered. The sequences are also
CC useful for isolating cDNAs and genes encoding homologous enzymes from the
CC same or other plant species, synthesising DNA primers and probes for
CC mapping the genes and for immunological screening of cDNA expression
CC libraries. The nucleic acid sequences can be used to manipulate the
CC isoflavone content without affecting other portions of the
CC phenylpropanoid pathway associated with lignin biosynthesis.

XX Sequence 327 AA;

Query Match 42.9%; Score 6; DB 21; Length 327;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDAGFE 9

Db 300 vdagfe 305

RESULT 9

AAB20313

ID AAB20313 standard; Protein; 543 AA.

XX AAB20313;

XX 29-MAY-2001 (first entry)

XX Soybean apoptosis inducing factor-like protein.

XX Apoptosis inducing factor; AIF; monodehydroascorbate reductase;
KW male sterility; stress response; tissue culture; plant; soybean.

XX Glycine max.

XX WO200119994-A2.

XX 22-MAR-2001.

XX 11-SEP-2000; 2000WO-US24859.

XX 13-SEP-1999; 99US-0153737.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Butler KH, Cahoon RE, Klein TM, Mazithulela G;

XX WPI; 2001-244804/25.

XX N-PSDB; AAF30442.

XX New polynucleotide encoding plant apoptosis inducing factor for
PT engineering male sterility, altering plant architecture, manipulating
PT stress response, and influencing growth of cells and tissues in culture

XX Claim 1(i); Page 71-72; 91pp; English.

XX The present sequence is that of a monodehydroascorbate reductase
CC like apoptosis inducing factor (AIF) of soybean, as predicted from
CC a full-length cDNA (see AAF30442). It shows 52.6% amino acid
CC identity to known monodehydroascorbate-reductase-like AIFs of Oryza
CC sativa and Brassica juncea. The invention relates to isolated
CC nucleic acids encoding plant AIFs. Such nucleic acids, used in
CC sense or antisense constructs, can be used to alter levels of AIFs
CC in plant cells, and hence to alter apoptosis and eventually to
CC control cell tissue culture growth, facilitate studies of programmed
CC cell death in plants, increase the efficiency of gene transfer, help
CC provide more stable transformations, engineer male sterility, alter
CC plant architecture and manipulate stress response.

XX Sequence 543 AA;

Query Match 42.9%; Score 6; DB 22; Length 543;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FEGQLT 13

Db 332 fegqlt 337

RESULT 10

AAP70143

ID AAP70143 standard; protein; 585 AA.

XX AAP70143;

XX 20-MAR-1991 (first entry)

XX BAR1 gene product.

XX BAR1 gene product; DNA construct; ss DNA; insulin; proinsulin;
XX Saccharomyces cerevisiae.

XX Key

XX Location/Qualifiers

XX Peptide

XX 1..24

XX /label=signal peptide

XX 84..90

XX /label=potential glycosylation site

XX 308..308

XX /label=potential glycosylation site

XX 366..366

XX /label=potential glycosylation site

```

FT FT Region 398..0 /label=potential glycosylation site
FT FT Region 468..0 /label=potential glycosylation site
FT FT Region 503..0 /label=potential glycosylation site
FT FT Region 551..0 /label=potential glycosylation site
XX PN EP220689-A.
XX XX
XX PD 06-MAY-1987.
XX PF 24-OCT-1986; 86EP-0114769.
XX PR 20-OCT-1986; 86WO-US02198.
XX PR 25-OCT-1985; 85US-0791305.
XX PA (MACK/) MACKAY V L.
XX PI Mackay VL;
XX XX
XX DR WPI: 1987-124280/18.
XX DR P-PSDB; AAP70143.
XX XX
XX XX New DNA fragment contg. yeast BAR I gene and foreign structural
XX PT gene - under control of promoter, esp. for producing insulin
XX PT which is secreted from the cell
XX PS Disclosure; Fig. 1A-C; 62pp; English.
XX XX
XX CC The BAR1 gene product is expressed as a fusion polypeptide/protein
XX CC in S.cerevisiae, esp. with insulin or proinsulin. The fusion protein
XX CC pref. also comprises a KEX 2 processing site. The presence of the BAR1
XX CC sequence means that the fusion protein is secreted into the culture
XX CC medium or periplasmic space, facilitating its purificn.
XX XX
XX SQ Sequence 585 AA;

Query Match 42.9%; Score 6; DB 8; Length 585;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTIVDA 6
Db 246 ptivda 251
|||||

RESULT 11
AAR13383
ID AAR13383 standard; Protein; 587 AA.
XX AC
XX AC AAR13383;
XX DT 29-OCT-1991 (first entry)
XX DE Barrier protein.
XX DE Heterologous protein; expression; secretion; urokinase; insulin;
XX KW EGF; TGF; PDGF.
XX OS Saccharomyces cerevisiae.
XX XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..24
XX FT Protein /label= sig_peptide
XX FT /label= mat_protein
XX FT Misc-difference 468
XX FT /label= ASN, GLN
XX FT Misc-difference 503
XX FT /label= ASN, GLN

FT FT Region 398..0 /label=potential glycosylation site
FT FT Region 468..0 /label=potential glycosylation site
FT FT Region 503..0 /label=potential glycosylation site
FT FT Region 551..0 /label=potential glycosylation site
XX PN EP220689-A.
XX XX
XX PD 06-MAY-1987.
XX PF 24-OCT-1986; 86EP-0114769.
XX PR 20-OCT-1986; 86WO-US02198.
XX PR 25-OCT-1985; 85US-0791305.
XX PA (MACK/) MACKAY V L.
XX PI Mackay VL;
XX XX
XX DR WPI: 1987-124280/18.
XX DR P-PSDB; AAP70143.
XX XX
XX XX New DNA fragment contg. yeast BAR I gene and foreign structural
XX PT gene - under control of promoter, esp. for producing insulin
XX PT which is secreted from the cell
XX PS Disclosure; Fig. 1A-C; 62pp; English.
XX XX
XX CC The BAR1 gene product is expressed as a fusion polypeptide/protein
XX CC in S.cerevisiae, esp. with insulin or proinsulin. The fusion protein
XX CC pref. also comprises a KEX 2 processing site. The presence of the BAR1
XX CC sequence means that the fusion protein is secreted into the culture
XX CC medium or periplasmic space, facilitating its purificn.
XX XX
XX SQ Sequence 585 AA;

Query Match 42.9%; Score 6; DB 8; Length 585;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTIVDA 6
Db 246 ptivda 251
|||||

RESULT 11
AAR13383
ID AAR13383 standard; Protein; 587 AA.
XX AC
XX AC AAR13383;
XX DT 29-OCT-1991 (first entry)
XX DE Barrier protein.
XX DE Heterologous protein; expression; secretion; urokinase; insulin;
XX KW EGF; TGF; PDGF.
XX OS Saccharomyces cerevisiae.
XX XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..24
XX FT Protein /label= sig_peptide
XX FT /label= mat_protein
XX FT Misc-difference 468
XX FT /label= ASN, GLN
XX FT Misc-difference 503
XX FT /label= ASN, GLN

FT FT Region 398..0 /label=potential glycosylation site
FT FT Region 468..0 /label=potential glycosylation site
FT FT Region 503..0 /label=potential glycosylation site
FT FT Region 551..0 /label=potential glycosylation site
XX PN US5037743-A.
XX PD 06-AUG-1991.
XX PF 14-NOV-1988; 88US-0270933.
XX PR 14-NOV-1988; 88US-0270933.
XX PR 02-OCT-1987; 87US-0104316.
XX PA (ZYMO-) ZYMOGENETICS INC.
XX PI Welch SK, Mackay VL, Yip CL;
XX DR WPI: 1991-252061/34.
XX DR N-PSDB; AAQ13195.
XX XX
XX PT DNA constructs for secretion of foreign proteins - using signal
XX PT sequence and portion of BAR1 C-terminal domain to direct secretion
XX PS Disclosure; Fig 1; 40pp; English.
XX XX
XX CC Asn468 and/or Asn503 may be replaced by Gln to prevent glycosylation.
XX CC A hybrid secretory peptide comprising the signal sequence and amino
XX CC acids 391-526 or 423-526 directs the secretion of heterologous
XX CC proteins or polypeptides, e.g. urokinase, insulin, platelet-derived
XX CC growth factor, epidermal growth factor or transforming growth factor
XX CC alpha.
XX CC See also AAQ13195-7.
XX XX
XX SQ Sequence 587 AA;

Query Match 42.9%; Score 6; DB 12; Length 587;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTIVDA 6
Db 248 ptivda 253
|||||

RESULT 12
AAR20109
ID AAR20109 standard; Protein; 587 AA.
XX AC
XX AC AAR20109;
XX DT 24-MAR-1992 (first entry)
XX DE BAR1 barrier protease.
XX DE Leucine-lysine cleavage; leucine-arginine cleavage.
XX OS Saccharomyces cerevisiae.
XX XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..24
XX FT /note= "signal peptide"
XX FT Peptide 25..587
XX FT /note= "mature peptide"
XX PN WO9118988-A.
XX PD 12-DEC-1991.
XX PF 05-JUN-1991; 91WO-US03952.
XX PR 24-SEP-1990; 90US-0587324.
XX PR 08-JUN-1990; 90US-0534933.
XX PA (ZYMO-) ZYMOGENETICS INC.

```


PI Mackay VL;
 DR WPI; 1992-007471/01.
 DR N-PSDB; AAQ20266.
 XX
 PT Pure barrier protease - useful in industrial processes where
 PT leucine-lysine or leucine-arginine cleavage is required at low pH
 PT or high temp.
 XX
 PS Disclosure; Fig 1; 42pp; English.
 XX
 CC The amino acid sequence is that of barrier protease (BP) from *S.*
 CC cerevisiae. BP compsns. may be used in a variety of industrial
 CC processes, partic. those calling for the cleavage of leucine-lysine
 CC or leucine-arginine bonds under acidic conditions or at high temp.
 CC BP may also be used in the prodn. and isolation of proteins made by
 CC genetic engineering methods, e.g. to cleave fusion proteins at
 CC Leu-Arg bonds.
 XX
 SQ Sequence 587 AA;
 Query Match 42.9%; Score 6; DB 13; Length 587;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PTIVDA 6
 DB 248 ptivda 253
 |||||
 RESULT 13
 AAY58677
 ID AAY58677 standard; Protein; 1001 AA.
 XX
 AC AAY58677;
 XX
 DT 25-APR-2000 (first entry)
 XX
 DE Arabidopsis RAN1 copper transporter.
 XX
 KW Copper transporter; RAN1; responsive-to-antagonist1;
 KW ethylene signalling pathway; ATPase; copper transport;
 KW transgenic plant.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200004760-A1.
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-US16591.
 XX
 PR 22-JUL-1998; 98US-0093698.
 XX
 PA (UYPE-) UNIV PENNSYLVANIA.
 XX
 PI Ecker JR, Hirayama T, Kieber JJ;
 DR WPI; 2000-171318/15.
 DR N-PSDB; AAZ57894, AAZ57895.
 XX
 PT New isolated plant copper transporter genes, useful for modulating e.g.
 PT germination, sex determination, flower or leaf senescence, fruit
 PT ripening, pathogen resistance or response to stress -
 XX
 PS Claim 7; Fig 5; 78pp; English.
 XX
 CC The present sequence is that of the RAN1 polypeptide of
 CC Arabidopsis thaliana ecotype Columbia (Col-0) as predicted from the
 CC exons of RAN1 cDNA (see AAZ57895). RAN1 is a copper transporting
 CC P-type ATPase that acts as an early acting regulator in the
 CC ethylene gas signalling pathway. Its shares structural features

CC with other copper transporters, including an N-terminal metal
 CC binding motifs, a phosphatase domain, a transduction domain, a
 CC phosphorylation domain, and an ATP binding domain. The invention
 CC relates to the ran1 gene and to antagonist responsive mutants
 CC controlling copper transport in the plant. It also relates to the
 CC manipulation of ran1 and its protein product to modulate the
 CC ethylene response in plants, thereby permitting the regulation and
 CC controlled alteration of plant growth and developmental processes,
 CC including germination, cell elongation, flower and leaf senescence,
 CC abscission, fruit ripening, insect, herbicide and pathogen
 CC resistance, and response to stress (claimed).
 XX
 SQ Sequence 1001 AA;
 Query Match 42.9%; Score 6; DB 21; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 DAGFEG 10
 DB 193 dagfeg 198
 |||||
 RESULT 14
 AAY58717
 ID AAY58717 standard; Protein; 1001 AA.
 XX
 AC AAY58717;
 XX
 DT 25-APR-2000 (first entry)
 XX
 DE Arabidopsis ran1-1 mutant protein.
 XX
 KW Copper transporter; ran1-1; responsive-to-antagonist1;
 KW ethylene signalling pathway; ATPase; copper transport;
 KW transgenic plant; mutant; mutein.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200004760-A1.
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-US16591.
 XX
 PR 22-JUL-1998; 98US-0093698.
 XX
 PA (UYPE-) UNIV PENNSYLVANIA.
 XX
 PI Ecker JR, Hirayama T, Kieber JJ;
 DR WPI; 2000-171318/15.
 DR N-PSDB; AAZ58168.
 XX
 PT New isolated plant copper transporter genes, useful for modulating e.g.
 PT germination, sex determination, flower or leaf senescence, fruit
 PT ripening, pathogen resistance or response to stress -
 XX
 PS Claim 30; Page -; 78pp; English.
 XX
 CC The present sequence is that of the Arabidopsis thaliana RAN1
 CC mutein ran1-1. The mutein has an amino acid substitution of Ile
 CC for Thr497 of the wild-type protein (see AAY58677). RAN1 is a
 CC copper transporter and early-acting regulator of the ethylene gas
 CC signalling pathway. Thr497 is in the phosphatase domain and is
 CC conserved in all copper transporters. The ran1-1 protein
 CC accumulates to the same level as RAN1 but has reduced copper
 CC transporting activity. The invention relates to the RAN1 gene and

CC to antagonist responsive mutants controlling copper transport in
CC the plant. It also relates to the manipulation of RAN1 and its
CC protein product to modulate the ethylene response in plants,
CC thereby permitting the regulation and controlled alteration of
CC plant growth and developmental processes, including germination,
CC cell elongation, flower and leaf senescence, abscission, fruit
CC ripening, insect, herbicide and pathogen resistance, and response
CC to stress (claimed).
CC Note: The present sequence is not shown in the specification but
CC is derived from the Arabidopsis RAN1 gene sequence given in
CC figure 5.
XX
SQ Sequence 1001 AA;

Query Match 42.9%; Score 6; DB 21; Length 1001;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DAGPEG 10
|||||
Db 193 dagfeg 198

RESULT 15
AAY58718
ID AAY58718 standard; Protein; 1001 AA.
AC AAY58718;
XX
XX 25-APR-2000 (first entry)
XX Arabidopsis ranl-2 mutant protein.
XX
XX Copper transporter; ranl-2; responsive-to-antagonist1;
XX ethylene signalling pathway; ATPase; copper transport;
XX transgenic plant; mutant; mutein.
XX
XX Arabidopsis thaliana.
XX
XX Key Location/Qualifiers
FH Misc-difference 173
FT /note= "replaces wild-type Gly"
XX
XX WO200004760-A1.
XX
XX 03-FEB-2000.
XX
XX 22-JUL-1999; 99WO-US16591.
XX
XX 22-JUL-1998; 98US-0093698.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Ecker JR, Hirayama T, Kieber JJ;
XX
XX WPI: 2000-171318/15.
XX N-PSDB; AA258168.
XX
XX New isolated plant copper transporter genes, useful for modulating e.g.
XX germination, sex determination, flower or leaf senescence, fruit
XX ripening, pathogen resistance or response to stress
XX
XX Claim 31; Page -: 78pp; English.
XX
XX The present sequence is that of the Arabidopsis thaliana RAN1
XX mutein ranl-2. The mutein has an amino acid substitution of Glu
XX for Gly173 of the wild-type protein (see AAY58677). RAN1 is a
XX copper transporter and early-acting regulator of the ethylene gas
XX signalling pathway. The ranl-2 protein accumulates to the same
XX level as RAN1 but has reduced copper transporting activity. The
XX invention relates to the RAN1 gene and to antagonist responsive
XX mutants controlling copper transport in the plant. It also relates

CC to the manipulation of RAN1 and its protein product to modulate the
CC ethylene response in plants, thereby permitting the regulation and
CC controlled alteration of plant growth and developmental processes,
CC including germination, cell elongation, flower and leaf senescence,
CC abscission, fruit ripening, insect, herbicide and pathogen
CC resistance, and response to stress (claimed).
CC Note: The present sequence is not shown in the specification but
CC is derived from the Arabidopsis RAN1 gene sequence given in
CC figure 5.
XX
SQ Sequence 1001 AA;

Query Match 42.9%; Score 6; DB 21; Length 1001;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DAGPEG 10
|||||
Db 193 dagfeg 198

Search completed: January 31, 2002, 13:18:11
Job time: 172 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:15 ; Search time 78.64 Seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-77

Perfect score: 14
Sequence: 1 AHRIDPGWSGCIVL 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	193	1 A42940	dCTP deaminase (EC
2	14	100.0	193	2 C85833	2'-deoxycytidine 5
3	8	57.1	195	1 A64050	dCTP deaminase (EC
4	8	57.1	206	2 C84942	dCTP deaminase (EC
5	6	42.9	141	2 E70110	hypothetical prote
6	6	42.9	300	2 H83385	probable transcrip
7	6	42.9	340	1 MMBE5	cell fusion protei
8	6	42.9	407	2 D85643	hypothetical prote
9	6	42.9	488	2 S71029	flak protein - Vib
10	6	42.9	733	2 S56277	probable membrane
11	6	42.9	980	2 T24336	hypothetical prote
12	6	42.9	2555	2 A40043	notch protein homo
13	5	35.7	78	1 A05008	translation initia
14	5	35.7	83	2 T35979	hypothetical prote
15	5	35.7	107	2 S38566	ferredoxin precurs
16	5	35.7	113	2 D86133	hypothetical prote
17	5	35.7	113	2 E82291	conserved hypothet
18	5	35.7	128	2 F71170	hypothetical prote
19	5	35.7	131	1 IEBC42	hypothetical prote
20	5	35.7	132	2 S56573	hypothetical 14.6k
21	5	35.7	141	2 B24338	hemoglobin alpha-pr
22	5	35.7	153	2 S04970	calcium-binding pr
23	5	35.7	193	2 T55544	Jag-related protei
24	5	35.7	197	2 T35766	hypothetical prote
25	5	35.7	198	2 T47962	hypothetical prote
26	5	35.7	229	2 E85806	hypothetical prote
27	5	35.7	230	2 S76965	hypothetical prote
28	5	35.7	249	2 T46836	precorrin-4 methyl
29	5	35.7	250	2 H75504	Mutr/nudix family

30 5 35.7 258 2 G72406
31 5 35.7 268 2 G70865
32 5 35.7 279 2 F84902
33 5 35.7 291 2 B82645
34 5 35.7 294 1 H64018
35 5 35.7 295 2 B85787
36 5 35.7 298 2 S73537
37 5 35.7 299 2 C70643
38 5 35.7 306 2 T39901
39 5 35.7 307 2 D83642
40 5 35.7 313 2 F96493
41 5 35.7 314 2 T34943
42 5 35.7 320 2 E71095
43 5 35.7 321 2 T41002
44 5 35.7 323 2 I64240
45 5 35.7 323 2 D82987

ALIGNMENTS

RESULT 1

A42940
dCTP deaminase (EC 3.5.4.13) dcd [validated] - Escherichia coli
N:Alternate names: dut (dUTPase) mutation suppressor
C:Species: Escherichia coli
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 26-May-2000
C:Accession: A42940; H64972
R:Wang, L.; Weiss, B.
J. Bacteriol. 174, 5647-5653, 1992
A:Title: dcd (dCTP deaminase) gene of Escherichia coli: mapping, cloning, sequencing,
A:Reference number: A42940; MUID:92380941
A:Accession: A42940
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-193 <W>
A:Cross-references: GB:M90069; NID:gl45715; PIDN:AAA23669.1; PID:gl45716
A:Note: sequence extracted from NCBI backbone (NCBIN:112699, NCBIP:112716)
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H64972
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-193 <BLAT>
A:Cross-references: GB:AE000296; GB:U00096; NID:gl788373; PIDN:AACT5126.1; PID:gl7883
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: dcd
C:Function:
A:Description: EC 3.5.4.13 [validated; MUID:92380941]
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHRIDPGWSGCIVL 14

Db 124 AHRIDPGWSGCIVL 137

RESULT 2

C85833
2'-deoxycytidine 5'-triphosphate deaminase [imported] - Escherichia coli (strain O157
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: C85833
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: C85833
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-193 <STO>
 A:Cross-references: GB:AE005174; NID:g12516264; PIDN:AAG57127.1; GSPDB:GN00145; UWGP:232
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: dcd
 C:Superfamily: dCTP deaminase

Query Match 100.0%; Score 14; DB 2; Length 193;
 Best Local Similarity 100.0%; Pred. No. 1.5e+09; Mismatches 0; Indels 0; Gaps 0;
 Matches 14; Conservative 0;

QY 1 AHRIDPGWGGCIVL 14
 |||||
 DB 124 AHRIDPGWGGCIVL 137

RESULT 3
 A64050
 dCTP deaminase (EC 3.5.4.13) H0133 [similarity] - Haemophilus influenzae (strain Rd KW2
 C:Species: Haemophilus influenzae
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
 C:Accession: A64050
 R:Feischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630
 A:Accession: A64050
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-195 <TIGR>
 A:Cross-references: GB:U32699; GB:L42023; NID:g3212180; PIDN:AAC21805.1; PID:g1573087;
 C:Superfamily: dCTP deaminase
 C:Keywords: hydrolase

Query Match 57.1%; Score 8; DB 1; Length 195;
 Best Local Similarity 100.0%; Pred. No. 0.018; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0;

QY 1 AHRIDPGW 8
 |||||
 DB 124 AHRIDPGW 131

RESULT 4
 C84942
 dCTP deaminase (EC 3.5.4.13) [imported] - Buchnera sp. (strain APS)
 N:Alternate names: deoxycytidine triphosphate deaminase
 C:Species: Buchnera sp.
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C:Accession: C84942
 R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H. Nature 407, 81-86, 2000
 A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
 A:Reference number: A84930; MUID:20445173
 A:Accession: C84942
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-206 <STO>
 A:Cross-references: GB:AP000398; GSPDB:GN00144
 A:Experimental source: strain APS
 C:Genetics:

A:Gene: dcd; BU108
 C:Keywords: hydrolase

Query Match 57.1%; Score 8; DB 2; Length 206;
 Best Local Similarity 100.0%; Pred. No. 0.019; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0;

QY 1 AHRIDPCW 8
 |||||
 DB 137 AHRIDPCW 144

RESULT 5
 E70110
 hypothetical protein BB0085 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 28-Jul-2000
 C:Accession: E70110
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: E70110
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-141 <SLE>
 A:Cross-references: GB:AE001121; GB:AE000783; NID:g2687963; PIDN:AAC66475.1; PID:g268
 A:Experimental source: strain B31
 C:Superfamily: Borrelia burgdorferi hypothetical protein BB0085

Query Match 42.9%; Score 6; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 3.3; Mismatches 0; Indels 0; Gaps 0;
 Matches 6; Conservative 0;

QY 9 SGCIVL 14
 |||||
 DB 58 SGCIVL 63

RESULT 6
 H83385
 probable transcription regulator PA2076 [imported] - Pseudomonas aeruginosa (strain P
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: H83385
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L ; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337
 A:Accession: H83385
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-300 <STO>
 A:Cross-references: GB:AE004635; GB:AE004091; NID:g9948085; PIDN:AAG05464.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2076

Query Match 42.9%; Score 6; DB 2; Length 300;
 Best Local Similarity 100.0%; Pred. No. 5.9; Mismatches 0; Indels 0; Gaps 0;
 Matches 6; Conservative 0;

QY 5 DPGWSG 10
 |||||
 DB 100 DPGWSG 105

```

R;Stewart, B.J.; McCarter, L.L.
Mol. Microbiol. 20, 137-149, 1996
A;Title: Vibrio parahaemolyticus FlaJ, a homologue of Flis, is required for productio
A;Reference number: S71027; MUID:97014377
A;Accession: S71029
A;Molecule type: DNA
A;Residues: 1-488 <STE>
A;Cross-references: EMBL:U12816; NID:g6806922; PIDN:AAC27806.1; PID:g1254206
A;Experimental source: strain BB22
C;Genetics:
A;Gene: flak
C;Superfamily: nitrogen assimilation regulatory protein ntrC; response regulator homo
F;137-358/Domain: RNA polymerase sigma factor interaction domain homology <SFI>

Query Match 42.9%; Score 6; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 WSGCIV 13
      |||||
Db 48 WSGCIV 53

RESULT 10
S56277
probable membrane protein YFR022w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 15-Sep-2000
C;Accession: S56277
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sas
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces
A;Reference number: S56186
A;Accession: S56277
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-733 <MUR>
A;Cross-references: EMBL:D50617; NID:g836685; PID:d1009902; PID:g836777; MIPS:YFR022w
C;Genetics:
A;Map position: 6R
C;Superfamily: Saccharomyces cerevisiae probable membrane protein YFR022w
C;Keywords: transmembrane protein
F;382-398/Domain: transmembrane #status predicted <TM>

Query Match 42.9%; Score 6; DB 2; Length 733;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGCIVL 14
      |||||
Db 43 SGCIVL 48

RESULT 11
T24336
hypothetical protein T01H3.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T24336
R;Burton, J.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19876
A;Accession: T24336
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-980 <NII>
A;Cross-references: EMBL:Z68317; PIDN:CRA92690.2; GSPDB:GN00020; CESP:T01H3.2
A;Experimental source: clone T01H3
C;Genetics:
A;Gene: CESP:T01H3.2
A;Map position: 2

cell fusion protein precursor - human herpesvirus 3
C;Species: human herpesvirus 3, varicella-zoster virus
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C;Accession: E27212
R;Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A;Title: The complete DNA sequence of varicella-zoster virus.
A;Reference number: A27345; MUID:86306657
A;Accession: E27212
A;Molecule type: DNA
A;Residues: 1-340 <DAV>
A;Cross-references: EMBL:X04370; NID:g59989; PIDN:CAA27888.1; PID:g59994
C;Genetics:
A;Gene: 5
C;Superfamily: herpesvirus cell fusion protein
C;Keywords: membrane fusion; transmembrane protein
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-340/Product: cell fusion protein #status predicted <CFP>
F;115-137/Domain: transmembrane #status predicted <TM1>
F;220-238/Domain: transmembrane #status predicted <TM3>
F;251-269/Domain: transmembrane #status predicted <TM4>
F;307-322/Domain: transmembrane #status predicted <TM5>

Query Match 42.9%; Score 6; DB 1; Length 340;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SGCIVL 14
      |||||
Db 226 SGCIVL 231

RESULT 8
D85643
hypothetical protein Z1487 [imported] - Escherichia coli (strain O157:H7)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C;Accession: D85643
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206351
A;Accession: D85643
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-407 <SPO>
A;Cross-references: GB:AE005174; NID:g12514343; PIDN:AAG55608.1; GSPDB:GN00145; UWGP:Z14
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1487

Query Match 42.9%; Score 6; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DPGWSG 10
      |||||
Db 117 DPGWSG 122

RESULT 9
S71029
flak protein - Vibrio parahaemolyticus
C;Species: Vibrio parahaemolyticus
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S71029

```

A: Introns: 25/1: 83/3: 242/1: 314/3: 357/3: 399/3: 437/3: 487/2: 561/3: 603/3: 636/3: 78
 C: Superfamily: Caenorhabditis elegans hypothetical protein t01H3.2

Query Match 42.9%; Score 6; DB 2; Length 980;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDPGWS 9
 |||||
 Db 116 IDPGWS 121

RESULT 12
 A40043
 notch protein homolog TAN-1 precursor - human
 C: Species: Homo sapiens (man)
 C: Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 13-Aug-1999
 C: Accession: A40043
 R: Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, Cell 66, 649-661, 1991
 A: Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal
 A: Reference number: A40043; MUID: 91347367
 A: Accession: A40043
 A: Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A: Molecule type: mRNA
 A: Residues: 1-2555 <ELL>
 A: Cross-references: GB: W73980
 F: Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F: 261-252/Domain: EGF homology <EGX1>
 F: 434-525/Domain: EGF homology <EGF1>
 F: 967-1018/Domain: EGF homology <EGX2>
 F: 1149-1180/Domain: EGF homology <EGF>
 F: 1187-1218/Domain: EGF homology <EGF>
 F: 1233-1264/Domain: EGF homology <EGX3>
 F: 1927-1959/Domain: ankyrin repeat homology <AN1>
 F: 1960-1992/Domain: ankyrin repeat homology <AN2>
 F: 1994-2026/Domain: ankyrin repeat homology <AN3>
 F: 2027-2059/Domain: ankyrin repeat homology <AN4>
 F: 2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 42.9%; Score 6; DB 2; Length 2555;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGWSG 10
 |||||
 Db 741 DPGWSG 746

RESULT 13
 A05008
 translation initiation factor IF-1 - liverwort (Marchantia polymorpha) chloroplast
 C: Species: Chloroplast Marchantia polymorpha
 C: Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
 C: Accession: A05008; S01562
 R: Ohyama, K.
 Submitted to the EMBL Data Library, October 1986
 A: Reference number: A00150
 A: Accession: A05008
 A: Molecule type: DNA
 A: Residues: 1-78 <OHY>
 A: Cross-references: EMBL: X04465; NID: g11640; PIDN: CAA28120.1; PID: g11709
 R: Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesono, K.; Shi Nature 322, 572-574, 1986
 A: Title: Chloroplast gene organization deduced from complete sequence of liverwort March
 A: Reference number: A38014
 A: Contents: Annotation; gene organization, sites, features
 R: Fukuzawa, H.; Kohchi, T.; Sano, T.; Shirai, H.; Umesono, K.; Inokuchi, H.; Ozeki, H.; J. Mol. Biol. 203, 333-351, 1988
 A: Title: Structure and organization of Marchantia polymorpha chloroplast genome. III. Ge
 A: Reference number: S01529; MUID: 89068687

A: Accession: S01562
 A: Molecule type: DNA
 A: Residues: 1-78 <FUK>
 A: Cross-references: EMBL: X04465; NID: g11640; PIDN: CAA28120.1; PID: g11709
 C: Genetics: inf
 A: Gene: inf
 A: Genome: chloroplast
 C: Superfamily: translation initiation factor IF-1
 C: Keywords: chloroplast; protein biosynthesis

Query Match 35.7%; Score 5; DB 1; Length 78;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GCIVL 14
 |||||
 Db 29 GCIVL 33

RESULT 14
 T35979
 hypothetical protein SC9C7.28 - Streptomyces coelicolor (fragment)
 C: Species: Streptomyces coelicolor
 C: Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C: Accession: T35979
 R: Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M submitted to the EMBL Data Library, January 1999
 A: Reference number: Z21551
 A: Accession: T35979
 A: Status: preliminary; translated from GB/EMBL/DDBJ
 A: Molecule type: DNA
 A: Residues: 1-83 <SEE>
 A: Cross-references: EMBL: AL035161; PIDN: CAA22740.1; GSPDB: GN00070; SC0EDB: SC9C7.28
 A: Experimental source: strain A3(2)
 C: Genetics:
 A: Gene: SC0EDB: SC9C7.28

Query Match 35.7%; Score 5; DB 2; Length 83;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHRID 5
 |||||
 Db 31 AHRID 35

RESULT 15
 S38566
 ferredoxin precursor - Psalteriomonas lanterna
 C: Species: Psalteriomonas lanterna
 C: Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
 C: Accession: S38566
 R: Brul, S.; Veltman, R.; Lombardo, M.C.P.; Vogels, G.D.
 Submitted to the EMBL Data Library, August 1993
 A: Description: Molecular cloning of ferredoxin cDNA from the anaerobic amoeboflagella
 A: Reference number: S38566
 A: Accession: S38566
 A: Molecule type: mRNA
 A: Residues: 1-107 <BRU>
 A: Cross-references: EMBL: X74556; NID: g414171; PIDN: CAA52650.1; PID: g414172
 A: Experimental source: strain Nijmegen
 C: Superfamily: Psalteriomonas lanterna ferredoxin
 F: 1-8/Domain: signal sequence #status predicted <SIG>
 F: 9-107/Product: ferredoxin #status predicted <MAT>

Query Match 35.7%; Score 5; DB 2; Length 107;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SGCIIV 13

Db |||||
 14 SGIV 18

Search completed: January 31, 2002, 13:20:16
Job time: 117 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:45 ; Search time 130.99 Seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-77
Perfect score: 14
Sequence: 1 AHRIDPGWGCIVL 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	42.9	81	5 Q9VBC3	Q9vbc3 drosophila
2	6	42.9	291	5 Q9NVJ2	Q9nvj2 haemochus
3	6	42.9	300	2 Q91239	Q91239 pseudomonas
4	6	42.9	422	2 Q9KXB5	Q9kxb5 escherichia
5	6	42.9	422	9 Q9XJL2	Q9xjl2 bacterioph
6	6	42.9	488	2 Q56708	Q56708 vibrio para
7	6	42.9	889	2 Q9AAZ6	Q9aaz6 caulobacter
8	6	42.9	980	5 Q22088	Q22088 caenorhabdi
9	6	42.9	2319	11 Q9R172	Q9r172 ratopus norv
10	5	35.7	83	2 Q9ZB12	Q9zb12 streptomyc
11	5	35.7	110	11 Q9CYF9	Q9cyf9 mus musculu
12	5	35.7	113	2 Q9KU41	Q9ku41 vibrio chol
13	5	35.7	128	1 Q58297	Q58297 pyrococcus
14	5	35.7	136	10 Q9LH77	Q9lh77 arabidopsis
15	5	35.7	138	10 Q9XG80	Q9xg80 oryza sativ
16	5	35.7	138	11 Q9D3Q2	Q9d3q2 mus musculu
17	5	35.7	143	4 Q9H935	Q9h935 homo sapien
18	5	35.7	148	2 Q9RPH3	Q9rph3 burkholderi
19	5	35.7	159	9 Q9MCA0	Q9mca0 bacterioph

20	5	35.7	161	10 Q9ASI3	Q9asi3 oryza sativ
21	5	35.7	163	11 Q9CXZ2	Q9cxz2 mus musculu
22	5	35.7	164	11 Q9CVP6	Q9cvp6 mus musculu
23	5	35.7	172	8 Q9TA01	Q9ta01 lampetra fl
24	5	35.7	193	2 Q9XR1	Q9xr1 deinococcus
25	5	35.7	195	2 Q69066	Q69066 pseudomonas
26	5	35.7	197	2 Q9S212	Q9s212 streptomyc
27	5	35.7	198	10 Q9M368	Q9m368 arabidopsis
28	5	35.7	203	11 Q9P5D9	Q9p5d9 mus musculu
29	5	35.7	217	4 Q9ULQ9	Q9ulq9 homo sapien
30	5	35.7	220	11 Q63404	Q63404 ratopus norv
31	5	35.7	220	12 Q88461	Q88461 stealth vir
32	5	35.7	223	10 Q9FSN2	Q9fsn2 oryza sativ
33	5	35.7	226	2 Q88116	Q88116 rhodobacter
34	5	35.7	228	11 Q9J113	Q9j113 rattus norv
35	5	35.7	228	11 Q9J105	Q9j105 mus musculu
36	5	35.7	228	11 Q9JHK0	Q9jhk0 mus musculu
37	5	35.7	230	2 P74757	P74757 synchocyst
38	5	35.7	233	12 Q65234	Q65234 african swi
39	5	35.7	233	12 Q65189	Q65189 african swi
40	5	35.7	236	5 Q9N4D9	Q9n4d9 caenorhabdi
41	5	35.7	244	2 Q9A8K0	Q9a8k0 caulobacter
42	5	35.7	247	10 Q9LN66	Q9ln66 arabidopsis
43	5	35.7	249	2 Q53138	Q53138 rhodococcus
44	5	35.7	250	2 Q9RWW5	Q9rww5 deinococcus
45	5	35.7	258	2 Q9WY34	Q9wy34 thermotoga

ALIGNMENTS

RESULT	ID	Q9VBC3	PRELIMINARY;	PRT;	81 AA.
AC	Q9VBC3				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DE	CG14244	PROTEIN.			
GN	CG14244				
OS	Drosophila melanogaster	(Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]	SEQUENCE FROM N.A.			
RP	STRAIN=BERKELEY;				
RX	MEDLINE=20196006;	PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,				
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,				
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,				

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.-W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AF003757; AAF56619.1; -.
DR FlyBase: FBgn0040607; CG14244.
SQ SEQUENCE 81 AA; 9185 MW; AF1E54DECDC8CE1 CRC64;

Query Match 42.9%; Score 6; DB 5; Length 81;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGWGGC 11
DB 76 PGWGGC 81

RESULT 2
Q9NJV2 PRELIMINARY; PRT; 291 AA.
AC Q9NJV2
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GALECTIN.
GN HCO-GAL-4.
OS Haemochus contortus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20183859; PubMed-10717307;
RA Greenhalgh C.J., Loukas A., Donald D., Nikolaou S., Newton S.E.;
RT "A family of galectins from haemonchus contortus.";
RL Mol. Biochem. Parasitol. 107:117-121(2000).
DR EMBL: AF105967; AAF63404.1; -.
DR InterPro: IPR001079; Gal-bind_lectin.
DR Pfam: PF00337; Gal-bind_lectin; 2.
DR SMART: SM00276; GLECT; 2.
DR PROSITE: PS00309; GALAPTIN; 2.
SQ SEQUENCE 291 AA; 33094 MW; DEAF6FF18088F8BC CRC64;

Query Match 42.9%; Score 6; DB 5; Length 291;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHRIDP 6
DB 264 AHRIDP 269

RESULT 3
Q9I239 PRELIMINARY; PRT; 300 AA.
AC Q9I239
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PROBABLE TRANSCRIPTIONAL REGULATOR.
GN PA2076.

```

```

OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=PA01;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brady L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL: AE004635; AG05464.1; -.
DR InterPro: IPR000847; HTH_LySR.
DR Pfam: PF00126; HTH_1; 1.
DR PRINTS: PR00039; HTHLYSR.
DR PROSITE: PS00044; HTH_LYSR FAMILY; UNKNOWN.1.
KW Complete proteome; DNA-binding; Transcription regulation.
SQ SEQUENCE 300 AA; 32458 MW; 21A9BF8F848BE9C CRC64;

Query Match 42.9%; Score 6; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPQWSG 10
DB 100 DPQWSG 105

RESULT 4
Q9KXB5 PRELIMINARY; PRT; 422 AA.
AC Q9KXB5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 45.5 KDA PROTEIN.
GN H0141.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7.
RX MEDLINE-20198780; PubMed-10734605;
RA Makino K., Yokoyama K., Kubota Y., Yutsudo C.H., Kimura S.,
RA Kurokawa K., Ishii K., Hattori M., Tatsuno I., Abe H., Iida T.,
RA Yamamoto K., Onishi M., Hayashi T., Yasunaga T., Honda T.,
RA Sasakawa C., Shinagawa H.;
RT "Complete nucleotide sequence of the prophage vt2-Sakai carrying the
RT verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
RL Genet. Syst. 74:227-239(1999).
DR EMBL: AP00422; BAA94169.1; -.
KW Hypothetical protein.
SQ SEQUENCE 422 AA; 45496 MW; 146BB1299BE985B0 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 422;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPQWSG 10
DB 132 DPQWSG 137

```

```

RESULT 5
O9XJL2 ID O9XJL2 PRELIMINARY; PRT; 422 AA.
AC O9XJL2
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 45.5 KDA PROTEIN.
GN L0125.
OS Bacteriophage 933W, and
OS Bacteriophage VT2-Sa.
OC Viruses; dsDNA viruses, no RNA stage; unclassified dsDNA phages.
OX NCBI_TaxID=10730, 97081;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-Phage 933W;
RX MEDLINE=99173898; PubMed=10074068;
RA Plunkett G. III, Rose D.J., Durfee T.J., Blattner F.R.;
RT "Sequence of Shiga toxin 2 phase 933W from Escherichia coli O157:H7;
RT Shiga toxin as a phage late-gene product.";
RL J. Bacteriol. 181:1767-1778(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-Phage 933W;
RA Plunkett G. III;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-Phage VT2-Sa;
RA Miyamoto H.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-Phage VT2-Sa;
RX MEDLINE=99419919; PubMed=10492170;
RA Miyamoto H., Nakai W., Yajima N., Fujibayashi A., Higuchi T., Sato K.,
RA Matsuhiro A.;
RT "Sequence analysis of Stx2-Converting phage VT2-Sa shows a great
RT divergence in early regulation and replication regions.";
RL DNA Res. 6:235-240(1999).
DR EMBL; AF125520; RAD25468.1; -.
DR EMBL; AP000363; BAA84345.1; -.
KW Hypothetical protein.
SQ SEQUENCE 422 AA; 45496 MW; 146BB1299BE985B0 CRC64;

Query Match 42.9%; Score 6; DB 9; Length 422;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGWSG 10
Db 132 DPGWSG 137

RESULT 6
O56708 ID O56708 PRELIMINARY; PRT; 488 AA.
AC O56708
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE POLAR FLAGELLAR PROTEIN.
GN FLAK.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB22;
RX MEDLINE=95189739; PubMed=7883718;
```

```

RA McCarter L.L.;
RT "Genetic and molecular characterization of the polar flagellum of
RT Vibrio parahaemolyticus.";
RL J. Bacteriol. 177:1595-1609(1995).
CC -I- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING
CC DOMAIN.
DR EMBL; AF069392; AAC27806.1; -.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR002197; HTH_Fis.
DR InterPro: IPR002078; Sig54_interact.
DR Pfam; PF00158; sigma54; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
DR PROSITE; PS00676; SIGMA54_INTERACT_2; 1.
DR PROSITE; PS00688; SIGMA54_INTERACT_3; 1.
DR PROSITE; PS00445; SIGMA54_INTERACT_4; 1.
KW ATP-binding; DNA-binding; Flagella; Transcription regulation.
SQ SEQUENCE 488 AA; 54330 MW; 25D560C091AA028 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 WSGCIV 13
Db 48 WSGCIV 53

RESULT 7
O9AAZ6 ID O9AAZ6 PRELIMINARY; PRT; 889 AA.
AC O9AAZ6
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TONB-DEPENDENT RECEPTOR.
GN CC0446.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.R., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005717; AAK22433.1; -.
DR TIGR; CC0446; -.
KW Receptor; Complete proteome.
SQ SEQUENCE 889 AA; 95775 MW; 75FCBD7A726A01A5 CRC64;
```

```

Query Match 42.9%; Score 6; DB 2; Length 889;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGWSG 10
Db 441 DPGWSG 446

RESULT 8
Q22088 ID Q22088 PRELIMINARY; PRT; 980 AA.
AC Q22088;
```

DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE T01H3.2 PROTEIN.
GN Caenorhabditis elegans.
OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Burton J.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z68317; CAA92690.2; -;
SQ SEQUENCE 980 AA; 111805 MW; BB69ACEAAD78D69C CRC64;

Query Match 42.9%; Score 6; DB 5; Length 980;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDPGWS 9
Db 116 IDPGWS 121

RESULT 9
Q9RL72 PRELIMINARY; PRT; 2319 AA.
ID Q9RL72
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NOTCH 3 PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Haritunians T., Boulter J., Weinmaster G., Schanen N.C.;
RT "Rattus norvegicus mRNA for Notch 3."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF164486; AAD46653.2; -;
DR HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF-Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; Ank; 6.
DR Pfam; PF00068; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PRO0010; EGFBL00D.
DR PRINTS; PRO0011; EGF_LAMININ.
DR PRINTS; PRO1452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 20.
DR SMART; SM00001; EGF-like; 12.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.

KW EGF-like domain; Glycoprotein; Hydroxylation.
SQ SEQUENCE 2319 AA; 244298 MW; 243BCA02D7C3283D CRC64;

Query Match 42.9%; Score 6; DB 11; Length 2319;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGWSG 10
Db 723 DPGWSG 728

RESULT 10
Q92BI2 PRELIMINARY; PRT; 83 AA.
ID Q92BI2
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE HYPOPHETICAL 9.4 KDA PROTEIN (FRAGMENT).
GN SC9C7.28
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
DR Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL035161; CAA22740.1; -;
KW Hypothetical protein.
FT NON_TER 83
SQ SEQUENCE 83 AA; 9361 MW; F13F85AE80A19DE6 CRC64;

Query Match 35.7%; Score 5; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHRID 5
Db 31 AHRID 35

RESULT 11
Q9CYF9 PRELIMINARY; PRT; 110 AA.
ID Q9CYF9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 5730493B19RIK PROTEIN.
GN 5730493B19RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

SQ SEQUENCE 136 AA; 15692 MW; 61916E1FF6E3316 CRC64;
 Query Match 35.7%; Score 5; DB 10; Length 136;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AHRID 5
 Db 89 AHRID 93
 RESULT 15
 Q9XG80 PRELIMINARY; PRT; 138 AA.
 AC Q9XG80;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PUTATIVE PHOSPHOLIPASE A2 PRECURSOR (EC 3.1.1.4).
 GN PLA2-I.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STPAIN-NIPPONBARE; TISSUE-GREEN SHOOT;
 RA Stahl U., Lee M., Sjoedahl S., Acher D., Cellini F., Ek B.,
 RA Inanacone R., Mackenzie D.A., Semeraro L., Tramontano E., Stymne S.;
 PT "Plant low molecular weight phospholipase A2s (PLA2s) are structurally
 PT related to the animal secretory PLA2s and are present as a family of
 PT isoforms in rice (Oryza sativa).";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ238116; CAB40841.1; -
 DR HSP; F14416; I8K9;
 DR InterPro; IPR001211; PLP_A2.
 DR ProDom; PD000303; PLP_A2; 1.
 DR PROSITE; PS00118; PA2_HIS; 1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 138 PUTATIVE PHOSPHOLIPASE A2.
 SQ SEQUENCE 138 AA; 14884 MW; CC0BC0F87A966C0B CRC64;
 Query Match 35.7%; Score 5; DB 10; Length 138;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GWSGC 11
 Db 52 GWSGC 56

Search completed: January 31, 2002, 13:37:47
 Job time: 172 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:13 ; Search time 65.13 Seconds
(without alignments)

4.837 Million cell updates/sec

Title: US-08-957-709-77

Perfect score: 14

Sequence: 1 AHRIDPGWGCIVL 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/2/iaa/PCFUS_COMB.pep.*
- 6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	US-08-822-774-58	Sequence 58, Appl
2	6	42.9	2556	US-08-185-432-17	Sequence 17, Appl
3	6	42.9	2556	US-08-083-590A-20	Sequence 20, Appl
4	6	42.9	2556	US-08-532-384-20	Sequence 20, Appl
5	5	35.7	20	US-09-400-208B-1	Sequence 1, Appl
6	5	35.7	32	US-08-361-920-9	Sequence 9, Appl
7	5	35.7	32	US-08-479-939-9	Sequence 9, Appl
8	5	35.7	32	US-08-483-432-9	Sequence 9, Appl
9	5	35.7	108	US-08-959-212-10	Sequence 10, Appl
10	5	35.7	145	US-08-997-080-55	Sequence 55, Appl
11	5	35.7	145	US-08-997-362-55	Sequence 55, Appl
12	5	35.7	145	US-08-873-970-55	Sequence 55, Appl
13	5	35.7	145	US-09-095-855-55	Sequence 55, Appl
14	5	35.7	145	US-08-705-347A-55	Sequence 55, Appl
15	5	35.7	191	US-08-766-605-3	Sequence 3, Appl
16	5	35.7	191	US-09-094-212-3	Sequence 3, Appl
17	5	35.7	201	US-08-933-750C-21	Sequence 21, Appl
18	5	35.7	201	US-09-234-613-21	Sequence 21, Appl
19	5	35.7	218	US-08-463-115-92	Sequence 92, Appl
20	5	35.7	218	US-08-465-388-92	Sequence 92, Appl
21	5	35.7	219	US-08-463-115-91	Sequence 91, Appl
22	5	35.7	219	US-08-465-388-91	Sequence 91, Appl
23	5	35.7	241	US-08-961-083-46	Sequence 46, Appl
24	5	35.7	264	US-08-463-115-93	Sequence 93, Appl
25	5	35.7	264	US-08-465-388-93	Sequence 93, Appl
26	5	35.7	273	US-08-997-080-75	Sequence 75, Appl
27	5	35.7	273	US-08-997-362-75	Sequence 75, Appl

28 5 35.7 273 3 US-08-873-970-75 Sequence 75, Appl
29 5 35.7 273 4 US-09-095-855-75 Sequence 75, Appl
30 5 35.7 284 1 US-08-411-777-10 Sequence 10, Appl
31 5 35.7 284 3 US-09-057-088-10 Sequence 10, Appl
32 5 35.7 286 4 US-09-254-733-3 Sequence 3, Appl
33 5 35.7 293 3 US-08-651-136C-20 Sequence 20, Appl
34 5 35.7 293 3 US-08-651-136C-24 Sequence 24, Appl
35 5 35.7 297 3 US-08-651-136C-4 Sequence 4, Appl
36 5 35.7 298 3 US-08-651-136C-18 Sequence 18, Appl
37 5 35.7 304 4 US-09-189-060B-72 Sequence 72, Appl
38 5 35.7 305 1 US-08-090-013-2 Sequence 2, Appl
39 5 35.7 305 1 US-08-081-328-2 Sequence 2, Appl
40 5 35.7 305 1 US-08-232-249-2 Sequence 2, Appl
41 5 35.7 305 2 US-08-921-426-8 Sequence 8, Appl
42 5 35.7 305 2 US-08-833-642A-2 Sequence 2, Appl
43 5 35.7 305 2 US-08-140-008A-4 Sequence 4, Appl
44 5 35.7 305 2 US-08-836-340-1 Sequence 1, Appl
45 5 35.7 305 2 US-08-389-423-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-08-822-774-58
; Sequence 58, Application US/08022774
; Patent No. 6183997

; GENERAL INFORMATION:

; APPLICANT: HOGREFE, Holly

; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)

; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,

; TITLE OF INVENTION: and Methods for Purifying and Identifying Same

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &

; ADDRESSEE: Lenahan, P.L.L.C.

; STREET: 1200 G Street, N.W. Suite 700

; CITY: Washington

; STATE: D.C.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/822,774

; FILING DATE: 21-MAR-1997

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: KULIK, David J.

; REGISTRATION NUMBER: 36,576

; REFERENCE/DOCKET NUMBER: 1486/43163

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 628-8800

; TELEFAX: (202) 628-8844

; INFORMATION FOR SEQ ID NO: 58:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 14 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-822-774-58

Query Match 100.0%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHRIDPGWGCIVL 14
|||||||

Db 1 AHRIDPGWGCIVL 14

RESULT 2

US-08-185-432-17
; Sequence 17, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Buseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEx PROTEINS, NUCLEIC ACIDS, AND
; REFERENCE/DOCKET NUMBER: 7326-006
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-17

Query Match 42.9%; Score 6; DB 1; Length 2556;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 DPGWSG 10

|||||

Db 742 DPGWSG 747

RESULT 3

US-08-083-590A-20
; Sequence 20, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; REFERENCE/DOCKET NUMBER: 7326-015
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York

COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-083-590A-20

Query Match 42.9%; Score 6; DB 1; Length 2556;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 DPGWSG 10

|||||

Db 742 DPGWSG 747

RESULT 4

US-08-532-384-20
; Sequence 20, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; REFERENCE/DOCKET NUMBER: 7326-015
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-20

Query Match 42.9%; Score 6; DB 3; Length 2556;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGWSG 10
DB 742 DPGWSG 747

RESULT 5
US-09-400-208B-1
Sequence 1, Application US/09400208B
Patent No. 6271011
GENERAL INFORMATION:
APPLICANT: Lee, Margie
APPLICANT: Sanchez, Susan
APPLICANT: Henk, Adam
TITLE OF INVENTION: Pasteurella Neureminidase Coding Sequences,
TITLE OF INVENTION: Compositions and Diagnostic Methods
FILE REFERENCE: 77-96B
CURRENT APPLICATION NUMBER: US/09/400.208B
PRIOR FILING DATE: 1999-09-21
PRIOR APPLICATION NUMBER: US 08/951,984
PRIOR FILING DATE: 1997-10-15
PRIOR APPLICATION NUMBER: US 60/028,876
PRIOR FILING DATE: 1996-10-16
PRIOR APPLICATION NUMBER: US 60/028,482
PRIOR FILING DATE: 1996-10-15
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 20
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: oligopeptide
OTHER INFORMATION: useful in immunogenic compositions
US-09-400-208B-1

Query Match 35.7%; Score 5; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RIDPG 7
DB 16 RIDPG 20

RESULT 6
US-08-361-920-9
Sequence 9, Application US/08361920
Patent No. 5457046
GENERAL INFORMATION:
APPLICANT: Woeldike, Helle F.
APPLICANT: Hagen, Frederick
APPLICANT: Hjort, Carsten M.
APPLICANT: Sven, Hastrup
TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
TITLE OF INVENTION: or Hemicellulose
NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 54570460 No. 5457046disk of No. 5457046th America, Inc.
STREET: 405 Lexington Avenue, 62nd Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/361.920
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/940,860
FILING DATE: 28-OCT-1992
APPLICATION NUMBER: DK 1158/90
FILING DATE: 09-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00124
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3435.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Humicola insolens
STRAIN: DSM 1800
US-08-361-920-9

Query Match 35.7%; Score 5; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GWSGC 11
DB 8 GWSGC 12

RESULT 7
US-08-479-939-9
Sequence 9, Application US/08479939
Patent No. 5686593
GENERAL INFORMATION:
APPLICANT: Woeldike, Helle F.
APPLICANT: Hagen, Frederick
APPLICANT: Hjort, Carsten M.
APPLICANT: Sven, Hastrup
TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
TITLE OF INVENTION: or Hemicellulose
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 56865930 No. 5686593disk of No. 5686593th America, Inc.
STREET: 405 Lexington Avenue, 62nd Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6201

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,939
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE: 22-DEC-1994
; APPLICATION NUMBER: US/07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Humicola insolens
; STRAIN: DSM 1800
; US-08-479-939-9

Query Match 35.7%; Score 5; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GWSGC 11
Db 8 GWSGC 12

RESULT 8
US-08-483-432-9
; Sequence 9, Application US/08483432
; Patent No. 5763254
; GENERAL INFORMATION:
; APPLICANT: Woeldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hott, Carsten M.
; APPLICANT: Svendsen, Hestrup
; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose
; TITLE OF INVENTION: On Hemicellulose
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5763254o No. 5763254disk of No. 5763254th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,432
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,920
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US/07/940,860
; FILING DATE: 28-OCT-1992
; APPLICATION NUMBER: DK 1158/90
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00124
; FILING DATE: 08-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3435.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Humicola insolens
; STRAIN: DSM 1800
; US-08-483-432-9

Query Match 35.7%; Score 5; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GWSGC 11
Db 8 GWSGC 12

RESULT 9
US-08-959-212-10
; Sequence 10, Application US/08959212
; Patent No. 6060274
; GENERAL INFORMATION:
; APPLICANT: Bjornvad, Mads
; APPLICANT: Schuelsen, Per
; APPLICANT: Jorgensen, Per
; TITLE OF INVENTION: Extracellular Expression Of Cellulose
; TITLE OF INVENTION: Binding Domains (CBD) Using Bacillus
; FILE REFERENCE: 4987.200-US
; CURRENT APPLICATION NUMBER: US/08/959,212
; CURRENT FILING DATE: 1997-10-28
; EARLIER APPLICATION NUMBER: 1192/96
; EARLIER FILING DATE: 1996-10-28
; EARLIER APPLICATION NUMBER: 1426/96
; EARLIER FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Humicola insolens
; US-08-959-212-10

Query Match 35.7%; Score 5; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 7 GWSG 11
11111
Db 84 GWSG 88

RESULT 10
US-08-997-080-55
; Sequence 55, Application US/08997080
; Patent No. 5968524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; CORRESPONDENCE ADDRESS: 194
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 37,007
; FILING DATE: June 12, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-997-080-55

Query Match 35.7%; Score 5; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHRID 5
11111
Db 117 AHRID 121

RESULT 11
US-08-997-362-55
; Sequence 55, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR

; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-997-362-55

Query Match 35.7%; Score 5; DB 2; Length 145;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHRID 5
11111
Db 117 AHRID 121

RESULT 12
US-08-873-970-55
; Sequence 55, Application US/08873970
; Patent No. 6001361
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: TREATMENT AND METHODS FOR
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICANT: Tan, Paul
FILING DATE: 29-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002C1
TELEPHONE: 206-269-0563
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-873-970-55

Query Match 35.7%; Score 5; DB 3; Length 145;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHRID 5
Db 117 AHRID 121

RESULT 13

US-09-095-855-55
Sequence 55, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002C3
TELEPHONE: 206-269-0563
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-095-855-55

Query Match 35.7%; Score 5; DB 4; Length 145;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHRID 5
Db 117 AHRID 121

RESULT 14

US-08-705-347A-55
Sequence 55, Application US/08705347A
Patent No. 6284255
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Speckman Picard PLLC
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/705.347A
FILING DATE: 28-AUG-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002
TELEPHONE: 206-269-0563
TELEFAX: 206-269-0563
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 145 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-705-347A-55

Query Match 35.7%; Score 5; DB 4; Length 145;

Best Local Similarity 100.0%; Pred. No. 59;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AHRID 5
|||||
Db 117 AHRID 121

RESULT 15
US-08-766-605-3
; Sequence 3, Application US/08766605
; Patent No. 5763220
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: HUMAN APOPTOSIS-RELATED CALCIUM-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,605
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0174 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 191 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1213520
US-08-766-605-3

Query Match 35.7%; Score 5; DB 1; Length 191;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 GCIVL 14
|||||
Db 154 GCIVL 158

Search completed: January 31, 2002, 13:15:14
Job time: 95 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:11 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-77

Perfect score: 14

Sequence: 1 AHRIDPGWGCIVL 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	14	100.0	14	19	AAW72853
2	6	42.9	53	20	AA114455
3	6	42.9	256	22	AA929766
4	6	42.9	733	22	AA929766
5	6	42.9	884	22	AA929766
6	5	35.7	10	22	AA929766
7	5	35.7	21	19	AA4869
8	5	35.7	32	12	AA15251
9	5	35.7	60	21	AA99905
10	5	35.7	60	22	AA00074
11	5	35.7	63	22	AA02803

12	5	35.7	70	22	AAE01617	Human gene 22 enco
13	5	35.7	95	21	AAE03260	Human secreted pro
14	5	35.7	101	21	AAE03790	Human secreted pro
15	5	35.7	102	21	AAE34611	Human secreted pro
16	5	35.7	108	19	AAE51350	Amino acid sequenc
17	5	35.7	117	20	AAE29194	Amino acid sequenc
18	5	35.7	124	21	AAE23792	Arabidopsis thalia
19	5	35.7	138	18	AAE35378	Protein encoded by
20	5	35.7	143	22	AAE95281	Human protein sequ
21	5	35.7	145	19	AAE60119	Mycobacterium vacc
22	5	35.7	145	20	AAE14865	M. vaccae antigen
23	5	35.7	154	21	AAE23791	Arabidopsis thalia
24	5	35.7	164	19	AAE60970	Streptococcus pneu
25	5	35.7	180	21	AAE63225	Gene 44 human secr
26	5	35.7	185	22	AAE93461	Human protein sequ
27	5	35.7	187	20	AAE29189	Amino acid sequenc
28	5	35.7	201	20	AAE93948	Human regulatory m
29	5	35.7	212	21	AAE58403	Lung cancer associ
30	5	35.7	218	19	AAE59198	Seq ID 92 from US
31	5	35.7	218	19	AAE40087	Seq ID 92 from US
32	5	35.7	219	19	AAE59197	Seq ID 91 from US
33	5	35.7	219	19	AAE40086	Seq ID 91 from US
34	5	35.7	241	19	AAE55085	Streptococcus pneu
35	5	35.7	242	22	AAE90786	C glutamicum prote
36	5	35.7	242	22	AAE76739	Corynebacterium gl
37	5	35.7	264	19	AAE59199	Seq ID 93 from US
38	5	35.7	264	19	AAE40088	Seq ID 93 from US
39	5	35.7	273	19	AAE60128	M. vaccae antigen
40	5	35.7	273	20	AAE14874	Partial amino acid
41	5	35.7	286	19	AAE57420	Humicola insolens
42	5	35.7	290	21	AAE07220	Arabidopsis thalia
43	5	35.7	290	21	AAE61558	Arabidopsis thalia
44	5	35.7	293	17	AAE04936	Chimeric endogluca
45	5	35.7	294	17	AAE04937	Chimeric endogluca

ALIGNMENTS

RESULT 1
AAW72853
ID AAW72853 standard; Peptide: 14 AA.
XX
AC AAW72853;
XX
DT 01-MAR-1999 (first entry)
XX
DE Escherichia coli dCTP deaminase uridine-binding motif.
DE Polymerase enhancing factor; PEF; dUTPase; PCR; amplification;
KW sequencing; replication.
XX
OS Escherichia coli.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
XX 20-MAR-1998; 98WO-US05497.
XX
XX 24-OCT-1997; 97US-0957709.
XX
XX 21-MAR-1997; 97US-0822774.
XX
XX (STRA-) STRATAGENE.
XX
XX Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes -
PT Improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication
XX

PS Claim 71; Page 47; 161pp; English.

XX This is the uridine-binding motif of the dCTP deaminase of

CC Escherichia coli. Sequences are provided (see AAW72849-57) of the

CC uridine-binding motifs of dUTPases and dCTP deaminases of

CC Pyrococcus furiosus (see AAW72847), Methanococcus jannaschii,

CC Desulfurolobus ambivalens, Escherichia coli, yeast, human and

CC herpesvirus; a consensus (see AAW72848) is also provided. A claimed

CC method of enhancing a nucleic acid polymerase reaction comprises

CC performing the reaction in the presence of one or more of the

CC following: a polymerase enhancing factor (PEF), a dUTPase, a

CC protein that turns-over dUTP and a protein having one or more of

CC the sequences provided in AAW72848-57. A claimed protein having PEF

CC activity comprises one or more of sequences given in AAW72848-57.

CC Kits are provided for replicating nucleic acids, for site-directed

CC mutagenesis, for nucleic acid sequencing or for amplification

CC (preferably PCR or RT-PCR).

XX Sequence 14 AA;

SQ

Query Match 100.0%; Score 14; DB 19; Length 14;

Best Local Similarity 100.0%; Pred. No. 3.7e-09;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHRIDPGWGCIVL 14

DB 1 ahridpgwsgcivl 14

RESULT 2

ID AAY14455 standard; Protein: 53 AA.

AC AAY14455;

XX 17-AUG-1999 (first entry)

DE Human secreted protein encoded by gene 45 clone HCFBJ91.

KW Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;

KW developmental abnormality; foetal deficiency; blood; allergy; renal;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;

KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

XX

XX WO9919339-A1.

XX 22-APR-1999.

XX 08-OCT-1998; 98WO-US21142.

XX 09-OCT-1997; 97US-0071498.

XX 09-OCT-1997; 97US-0061463.

XX 09-OCT-1997; 97US-0061527.

XX 09-OCT-1997; 97US-0061529.

XX 09-OCT-1997; 97US-0061532.

XX 09-OCT-1997; 97US-0061536.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Brewer LA, Duan R, Ebner R, Ferrie AM, Florence C;

PI Florence KA, Greene JM, Olsen HS, Rosen CA, Ruben SM;

PI Young PE, Yu G;

XX WPI; 1999-277587/23.

DR N-PSDB; AAX79055.

XX

PS Claim 11; Page 203; 226pp; English.

XX This sequence represents a secreted human protein encoded by the gene

CC clone detailed in the descriptor line. The gene can be used to generate

CC fusion proteins by linking to the gene to a human immunoglobulin Fc

CC portion (e.g. AAX79002) for increasing the stability of the fused

CC protein as compared to the human protein only.

CC The invention relates to 53 novel genes and their fragments (nucleic

CC acid sequences: AAX79011-X79064; amino acid sequences AAY14411-Y14464)

CC which are useful for preventing, treating or ameliorating medical

CC conditions e.g. by protein or gene therapy. Also, pathological

CC polypeptides in a sample or by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for each of the 53

CC polynucleotides, based on which tissues they are most highly expressed in

CC (see AAX79011 for described uses).

XX Sequence 53 AA;

SQ

Query Match 42.9%; Score 6; DB 20; Length 53;

Best Local Similarity 100.0%; Pred. No. 5.9;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RIDPGW 8

DB 41 ridpgw 46

RESULT 3

ID AAG92766 standard; Protein: 256 AA.

AC AAG92766;

XX 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6520.

KW Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

DR N-PSDB; AAG67985.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX

PS Claim 17; SEQ ID NO: 6520; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 256 AA;

Query Match 42.9%; Score 6; DB 22; Length 256;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 IDPGWS 9
Db 223 idpgws 228
|||||

RESULT 4
AAG70721
ID AAG70721 standard; Protein; 733 AA.
XX
AC AAG70721;
XX
DT 27-JUL-2001 (first entry)
XX
DE S cerevisiae apoptosis associated protein YFL015C.
XX
KW Yeast; fungus; apoptosis; infection; proliferative disease;
KW vaccine; autoimmune disease; ischaemia; neurodegeneration.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200102550-A2.
XX
PD 11-JAN-2001.

XX
PF 03-JUL-2000; 2000WO-BE00077.
XX
PR 01-JUL-1999; 99EP-0870141.
XX
PA (JANC) JANSSEN PHARM NV.

XX Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;
PI Nelissen BJM, Reekmans RJ;
XX
DR WPI: 2001-367042/38.
DR N-PSDB; AAH29757.

XX Yeast and fungal nucleic acids encoding proteins involved in a pathway
XX leading to programmed cell death, useful for treating proliferative
XX disorders, yeast and fungal infections, or for preventing apoptosis in
XX certain diseases -

PS Claim 1; Fig 1; 218pp; English.

XX The present invention provides the protein and coding sequences of a
XX number of apoptosis associated proteins from the yeast Saccharomyces
XX cerevisiae and the fungus Candida albicans. These can be used to identify
XX treatments for fungal and yeast infections, for proliferative diseases
XX and for apoptosis related diseases such as autoimmune diseases, ischaemia
XX and neurodegeneration. The present sequence is one of the S. cerevisiae
XX proteins of the invention.

SQ Sequence 733 AA;

Query Match 42.9%; Score 6; DB 22; Length 884;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 42.9%; Score 6; DB 22; Length 733;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 SGCIVL 14
Db 43 sgci vl 48
|||||

RESULT 5
AAU03556
ID AAU03556 standard; Protein; 884 AA.

XX
AC AAU03556;

XX
DT 12-SEP-2001 (first entry)

XX
DE Human protein kinase #56.

XX
KW Human; protein kinase; PTK; STK; cancer; cardiovascular disease;
KW metabolic disorder; immune related disease; neurological disorder;
KW neurodegenerative disorder; inflammatory disorder; infectious disease;
KW reproductive disorder.

XX
OS Homo sapiens.

XX
PN WO200138503-A2.

XX
PD 31-MAY-2001.

XX
PF 22-NOV-2000; 2000WO-US32085.

XX
PR 24-NOV-1999; 99US-0167482.

XX
PA (SUGE-) SUGEN INC.

XX
PI Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
PI Flanagan P, Clary D;

XX
DR WPI: 2001-343950/36.
DR N-PSDB; AAS06756.

XX Nucleic acids encoding human kinase polypeptides, useful for preventing
XX diagnosing and/or treating e.g. cancer, immune, cardiovascular and
XX neuronal-associated diseases, and microbial infections -

PS Claim 7; Figure 2; 433pp; English.

XX AAU03501-AAU03557 represent novel human protein kinases #1-57. The
XX novel protein kinases have been identified as members of the tyrosine
XX or serine/threonine kinase (PTK and STK) families. The polynucleotides
XX encoding protein kinases and the polypeptides may be used in the
XX prevention, diagnosis and treatment of diseases associated with
XX inappropriate kinase expression. For example, they may be used to treat
XX cancers (especially cancers of haematopoietic origin), cardiovascular
XX disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
XX immune related diseases (e.g. rheumatoid arthritis), neurological
XX disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
XX Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
XX disease (e.g. HIV) and reproductive disorders (e.g. infertility).
XX Additionally, polynucleotides encoding protein kinases may be
XX used for gene therapy and as DNA probes in diagnostic assays.

XX The protein kinase polypeptides may be used as antigens in the production
XX of antibodies against the protein kinases and in assays to identify
XX modulators of protein kinase expression and activity.

SQ Sequence 884 AA;

```

Oy 5 DPGWSG 10
Db 320 dpgwsg 325

RESULT 6
AAG96758
ID AAG96758 standard; Peptide: 10 AA.
XX
AC AAG96758;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 2952.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB04776.
XX
PR 13-DEC-1999; 99GB-0029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
DR WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides
PT to proteins encoded by genes of the human genome, useful in an assay
PT for screening and identifying of one or more novel peptides which are
PT drug candidates or pro-drugs.
XX
PS Example 4; Page 466; 646pp: English.
XX
CC The invention relates to a set of complementary peptide ligands
CC generated from the human genome. The complementary peptides
CC interact with their relevant target proteins encoded in the human
CC genome. They can be used as reagents in drug discovery and as lead
CC ligands to facilitate drug design and development. The present
CC sequence is a complementary peptide provided in the specification.
XX
SQ Sequence 10 AA;

Query Match 35.7%; Score 5; DB 22; Length 10;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 DPGWSG 10
Db 4 dpgwsg 8

RESULT 7
AAW48869
ID AAW48869 standard; peptide; 21 AA.
XX
AC AAW48869;
XX
DT 15-SEP-1998 (first entry)
XX
DE Pasteurella multocida neuraminidase immunogenic peptide.
XX
KW Pasteurella multocida strain R1913 NanH; neuraminidase; detection;
KW diagnosis; fowl cholera; shipping fever; respiratory tract infection;
KW abscess; systemic infection; immunoassay; immunogenic.
XX
XX

Pasteurella multocida.
XX
WO9816649-A1.
XX
23-APR-1998.
XX
15-OCT-1997; 97WO-US18668.
XX
16-OCT-1996; 96US-0028876.
XX
15-OCT-1996; 96US-0028482.
XX
(UYGE-) UNIV GEORGIA RES FOUND INC.
XX
Henk A, Lee MD;
XX
WPI; 1998-271747/24.
XX
Pasteurella multocida neuraminidase - useful for protecting animals
and humans from P. multocida infections
XX
Claim 6; Page 26; 61pp; English.
XX
The present sequence represents a Pasteurella multocida strain R1913
NanH neuraminidase immunogenic peptide. The present invention also
describes an immunogenic composition comprising a recombinant
neuraminidase from P. multocida (shown in AAW48868) or the immunogenic
peptides shown in AAW48869 to AAW48871; and an immunogenic assay kit for
the diagnosis and/or detection of P. multocida, comprising an antibody
specific to P. multocida neuraminidase and reagents for detecting the
specific binding of the neuraminidase and antibody. The neuraminidase
protein and immunogenic peptides can be used to immunise and protect
animals, e.g. sheep, cattle, rabbits, dogs, cats, chickens, turkeys and
humans against P. multocida diseases, such as fowl cholera, shipping
fever in cattle, respiratory tract infections, abscesses and systemic
infections in various animals. The nucleic acid sequence and systemic
oligonucleotides derived from it, or the antibody can be used to
diagnose P. multocida infections.
XX
SQ Sequence 21 AA;

Query Match 35.7%; Score 5; DB 19; Length 21;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 RIDPG 7
Db 17 ridpg 21

RESULT 8
AAR15251
ID AAR15251 standard; Protein; 32 AA.
XX
AC AAR15251;
XX
DT 12-FEB-1992 (first entry)
XX
DE Carbohydrate binding domain #8.
XX
KW cellulose; CBD; hemicellulosic substrate;
KW Trichoderma reesei; cellulase; terminal A region.
XX
WO9117244-A.
XX
14-NOV-1991.
XX
08-MAY-1991; 91WO-DK00124.
XX
09-MAY-1990; 90DK-0001158.
XX
(NOVO ) NOVO NORDISK A/S.
XX
XX

```

PI Woldlike HF, Hagen F, Hjort CM, Hastrup S;
XX WPI; 1991-353766/48.
XX
PT New fungal (hemi)cellulose degrading enzymes - for prodn. of liq.
PT fuel gas and feed protein, have specified carbohydrate binding domain
XX
PS Claim 20; Page 45; 73pp; English.
XX
CC This CBD is homologous to a terminal A region of *Trichoderma reesei*
CC cellulases and effects binding of a protein to an insoluble
CC cellulosic or hemicellulosic substrate. It is one of ten specific
CC CBD's (see AAR15244-RI5253) which correspond to the generic CBD
CC formulae in AAR15242 and AAR15243. The CBD is incorporated into a fusion
CC protein comprising a catalytic domain from a cellulase, e.g. a
CC *Bacillus endoglucanase*, and optionally comprising a linking B domain
CC from e.g. a fungal endoglucanase.
XX
XX Sequence 32 AA;
SQ

Query Match 35.7%; Score 5; DB 12; Length 32;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GWSGC 11
DB 8 gwsGC 12
|||||

RESULT 9
AAY99905
ID AAY99905 standard; Peptide; 60 AA.
XX
AC AAY99905;
XX
DT 16-NOV-2000 (first entry)
XX
DE Peptide encoded by Snut 1700Eagl DNA used in HIV DNA vaccine.
XX
KW HIV; human immunodeficiency virus; vaccine; AIDS; snut;
KW silent nucleotide substitution.
XX
OS Human immunodeficiency virus type 1.
XX
PN WO200029561-A2.
XX
PD 25-MAY-2000.
XX
PF 27-MAR-2000; 2000WO-DK00144.
XX
PR 29-MAR-1999; 99DK-0000427.
PR 09-APR-1999; 99US-0128558.
XX
PA (STAT-) STATENS SERUM INST.
XX
PI Pomsgaard A;
XX
DR WPI; 2000-387778/33.
DR N-PSDB; AAA49074.
XX
XX Producing nucleotide sequence construct with optimized codons for human
PT immunodeficiency virus (HIV) genetic vaccine involves obtaining a first
PT nucleotide sequence from a HIV patient, redesigning and assembling it
PT with snuts -
XX
PS Example 3; 104; 150pp; English.
XX
CC The present invention relates to a nucleotide construct with optimised
CC codons for use as a human immunodeficiency virus (HIV) DNA vaccine. The
CC construct uses codons from highly expressed mammalian proteins to code
CC for each derivative of an early, primary HIV envelope gene. The first
CC stage in the production of the construct was the cloning of an HIV

CC envelope gene. A nucleotide sequence encoding this gene was then created
CC using codons from highly expressed mammalian genes. The present sequence
CC is the peptide encoded by one of the snuts (AAA49060-A49079) that were
CC created by redesigning the nucleotide construct so that restriction
CC enzyme sites surrounded functional regions of the sequence. The snuts
CC were then assembled into pieces (AAA49080-A49092). Each derivative of
CC the envelope gene (AAA49093-A49097) was then built using the pieces.
CC The HIV DNA vaccine may be used as a prophylactic vaccine and as a
CC therapeutic vaccine in HIV infected patients.
XX
XX Sequence 60 AA;
SQ

Query Match 35.7%; Score 5; DB 21; Length 60;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGWSG 10
DB 20 pgwsg 24
|||||

RESULT 10
AAM00074
ID AAM00074 standard; Protein; 60 AA.
XX
AC AAM00074;
XX
DT 28-SEP-2001 (first entry)
XX
DE 3-Deoxy-D-arbino-heptulosonate 7-phosphate synthase sequence #142.
XX
KW Moss; carbohydrate metabolism related protein; CMPR; sugar; cofactor;
KW fine chemical production; carbohydrate; polysaccharide.
XX
OS Physcomitrella patens.
XX
PN WO200144476-A2.
XX
PD 21-JUN-2001.
XX
PF 14-DEC-2000; 2000WO-EPI2697.
XX
PR 16-DEC-1999; 99US-0171101.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;
PI Frank M, Freund A, Duwenig E, Schmidt R, Reski R;
XX
XX WPI; 2001-398155/42.
DR N-PSDB; AAH88760.
XX
PT Novel moss nucleic acid molecules encoding a carbohydrate metabolism
PT related protein useful for modulating production of fine chemicals such
PT as carbohydrates, cofactors and enzymes from microorganisms and plants
PT
XX
PS Claim 30; Page 122; 133pp; English.
XX
XX This invention relates to nucleic acid molecules AAH88708 - AAH88796
CC isolated from *Physcomitrella patens* (a moss), which encode carbohydrate
CC metabolism related proteins (CMPR) represented in AAM00022 - AAM00110.
CC Included in the invention is a vector containing the CMPR cDNA, and a
CC host cell transformed with the vector. The host cell (a microorganism,
CC *Corynebacterium* or *Brevibacterium*, moss or algae or a plant cell) is
CC useful for producing a fine chemical such as carbohydrates, cofactors
CC and/or enzymes. The nucleic acid molecules are suitable for modifying a
CC carbohydrate production system in a host, e.g., microorganisms and
CC plants. They are also useful to identify those DNA sequences and enzymes
CC in other species which are useful to modify the biosynthesis of starch,
CC cell wall polysaccharides and sugars. The nucleic acid molecules may be
CC utilised in the genetic engineering of *Corynebacterium glutamicum* and the

CC related Brevibacterium species and Acetobacter xylinum and Chlorella to
 CC make it a better or more efficient producer of one or more fine
 CC chemicals. Mutagenesis of one or more CMRPs may also result in CMRPs
 CC having altered activities which indirectly impact the production of one
 CC or more desired fine chemicals from plants. Primers AAH8705 - AAH80707
 CC are used in the sequencing of the CMRP cDNA sequences of the invention.
 XX
 SQ Sequence 60 AA;

Query Match 35.7%; Score 5; DB 22; Length 60;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHRID 5
 Db 5 ahrid 9
 |||||

RESULT 11
 AAU02803 standard; Protein: 63 AA.
 ID AAU02803
 AC AAU02803;
 DT 07-SEP-2001 (first entry)
 DE Taxus cuspidata RT-PCR generated amplicon polypeptide #4.
 XX
 OS Oxygenase; Japanese yew; Taxol; taxoid; Taxol biosynthetic pathway;
 KW transgenic organism.
 KW
 XX Taxus cuspidata.
 OS
 XX WO200134780-A2.
 PN
 XX 17-MAY-2001.
 PD
 XX 13-NOV-2000; 2000WO-US31254.
 XX
 PF 12-NOV-1999; 99US-0165250.
 PR
 XX (UNIW) UNIV WASHINGTON STATE RES FOUND.
 PA
 XX Croteau RB, Schoendorf A, Jennewein S;
 PI WPI: 2001-355489/37.
 DR N-PSDB; AAS05148.

XX Novel nucleic acid and amino acid sequences, isolated from the Taxus
 PT genus, useful for the synthetic production of taxol and related
 PT taxoids, intermediates within the Taxol biosynthetic pathway, and
 PT other taxoid derivatives -
 XX

PS Claim 1; Page 91; 141pp; English.

XX The sequence represents a Taxus cuspidata oxygenase, encoded by a DNA of
 CC the invention. The sequences isolated from the taxus genus, and the
 CC respective oxygenases are useful for the synthetic production of Taxol
 CC and related taxoids, as well as intermediates within the taxol
 CC biosynthetic pathway, and other taxoid derivatives. The sequences also
 CC can be used to make transgenic organisms that either produce the
 CC oxygenases for subsequent in vitro use, or produce the oxygenases in vivo
 CC so as to alter the level of taxol and taxoid production within the
 CC transgenic organism. The oxygenase nucleic acids and amino acids are
 CC useful for isolating the polynucleotide and polypeptide sequences
 CC corresponding to full-length oxygenases.

XX Sequence 63 AA;

Query Match 35.7%; Score 5; DB 22; Length 63;
 Best Local Similarity 100.0%; Pred. No. 84;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 SGCIV 13
 Db 31 sgclv 35
 |||||

RESULT 12
 AAEO1617 standard; Protein: 70 AA.
 ID AAEO1617
 AC AAEO1617;
 DT 17-JUL-2001 (first entry)
 DE Human gene 22 encoded secreted protein HT4ES80, SEQ ID NO:167.
 XX
 OS Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; aschma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angioinetic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnerability;
 KW cell culture; chemotaxis; food additive; gene therapy;
 KW binding partner identification; chromosome 16.
 XX

OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1
 FT /label= Signal_peptide
 FT Protein 2..70
 FT /label= Mature_human_secreted_protein
 FT Misc-difference 20
 FT /label= Unknown
 FT /note= "Encoded by WTC"
 FT Misc-difference 60
 FT /label= Unknown
 FT /note= "Encoded by TWA"
 XX
 PN WO200134623-A1.
 XX 17-MAY-2001.
 PD
 XX 01-NOV-2000; 2000WO-US30037.
 PF
 XX 05-NOV-1999; 99US-0163577.
 PR 30-JUN-2000; 2000US-0215137.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Ruben SM, Komatsoulis GA, Moore PA;

XX WPI: 2001-316490/33.
 DR N-PSDB; AAD05460.

XX Nucleic acids encoding 29 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX

PS Claim 11; Page 505; 535pp; English.

XX AAD05389-AAD05473 represent cDNAs corresponding to 29 human secreted
 CC protein genes, and AAEO1546-AAEO1630 represent the proteins they encode.
 CC AAEO1631-AAEO1660 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the

CC 29 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein of the invention.

XX Sequence 70 AA;

Query Match 35.7%; Score 5; DB 22; Length 70;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GWSGC 11
 Db 32 gwsqc 36
 |||||

RESULT 13
 AAG03260

ID AAG03260 standard; Protein; 95 AA.

XX AAG03260;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 7341.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX N-PSDB; AAC03266.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 7341; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30

CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 95 AA;

Query Match 35.7%; Score 5; DB 21; Length 95;
 Best Local Similarity 100.0%; Pred. No. 11e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGWSG 10
 Db 29 pgwsg 33
 |||||

RESULT 14

AAG03790

ID AAG03790 standard; Protein; 101 AA.

XX AAG03790;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 7871.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX N-PSDB; AAC03796.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 7871; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

SQ Sequence 101 AA;

Query Match 35.7%; Score 5; DB 21; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGWSG 10
|||||

Db 82 pgwsg 86

RESULT 15

AAB34611

ID AAB34611 standard; Protein; 102 AA.

AC AAB34611;

DT 26-JAN-2001 (first entry)

DE Human secreted protein sequence encoded by gene 35 SEQ ID NO:95.

KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; cancer; cardiovascular disorder; infection;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW ocular disorder; wound healing; skin aging; food additive; preservative.

XX Homo sapiens.

XX WO200056751-A1.

XX 28-SEP-2000.

XX 09-MAR-2000; 2000WO-US06013.

XX 19-MAR-1999; 99US-0125360.

XX 11-JUN-1999; 99US-0138626.

XX 03-DEC-1999; 99US-0168662.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-579482/54.

XX N-PSDB; AAC59772.

XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -

XX Claim 11; Page 384; 419pp; English.

XX The polynucleotide sequences given in AAC59738 to AAC59787 encode the
XX human secreted proteins given in AAB34577 to AAB34626. AAB34627 to
XX AAB34686 represent human secreted polypeptide sequences and proteins
XX homologous to them, which are given in the exemplification of the present
XX invention. Human secreted proteins have activities based on the tissues
XX and cells the genes are expressed in. Example of activities include:
XX antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
XX cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;
XX neuroprotective; antibacterial; virucide; fungicide; and
XX ophthalmological. The polynucleotides and proteins can be used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also
XX used in diagnosing a pathological condition or susceptibility to a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases, hyperproliferative disorders e.g. neoplasms and
XX cancers of the breast or liver, cardiovascular disorders,
XX cerebrovascular disorders, angiogenesis, nervous system disorders,
XX infections caused by bacteria, viruses and fungi and ocular disorders.
XX The proteins can also be used to aid wound healing and epithelial cell

CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues and in chemotaxis. The proteins can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59729 to AAC59737 and AAC34576 represent sequences used
CC in the exemplification of the present invention.

XX Sequence 102 AA;

Query Match 35.7%; Score 5; DB 21; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGWSG 10

Db 54 pgwsg 58

Search completed: January 31, 2002, 13:18:12
Job time: 173 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:47 ; Search time 130.99 Seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-78

Perfect score: 14

Sequence: 1 VGLIDSDYQGQLMI 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 14627329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_invertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	71.4	139	2 Q9F7S4	Q9f7s4 uncultured
2	10	71.4	150	2 Q9JZU7	Q9jzu7 neisseria m
3	10	71.4	150	2 Q9JUM1	Q9juw1 neisseria m
4	10	71.4	151	2 Q9HTN3	Q9htn3 pseudomonas
5	9	64.3	155	2 Q9PGZ6	Q9pgz6 xylella fas
6	6	42.9	139	10 Q9LSA1	Q9lsal arabidopsis
7	6	42.9	232	9 O03945	O03945 bacterioph
8	6	42.9	279	5 Q9U8P7	Q9u8p7 dictyosteli
9	6	42.9	280	2 P95727	P95727 streptomyce
10	6	42.9	355	2 Q9RE73	Q9rp73 pasteurella
11	6	42.9	388	2 Q99X85	Q99x85 staphylococ
12	6	42.9	404	1 Q29700	Q29700 archaeoglob
13	6	42.9	433	5 Q9NKA0	Q9nka0 drosophila
14	6	42.9	438	5 Q9VJQ3	Q9vjg3 drosophila
15	6	42.9	455	3 Q9Y7W5	Q9y7w5 saccharomyc
16	6	42.9	464	10 Q9SGA8	Q9sga8 arabidopsis
17	6	42.9	517	3 Q9C1Z8	Q9c1z8 pichia past
18	6	42.9	586	3 Q9URE1	Q9ure1 saccharomyc
19	6	42.9	586	3 Q9P986	Q9p986 saccharomyc

20 20 6 42.9 586 3 Q9P985
21 21 6 42.9 586 3 Q9P984
22 22 6 42.9 650 2 Q31382
23 23 6 42.9 724 10 Q9ZT83
24 24 6 42.9 847 2 Q9X9F5
25 25 6 42.9 1142 3 Q9P797
26 26 6 42.9 1148 5 Q17517
27 27 6 42.9 1214 2 O50463
28 28 6 42.9 1260 2 Q9CC97
29 29 6 42.9 1272 2 Q9FBR4
30 30 6 42.9 1540 4 Q9P111
31 31 6 42.9 2052 6 P79114
32 32 6 42.9 2058 4 Q9NYM7
33 33 6 42.9 2058 4 Q9HD67
34 34 6 42.9 2062 11 Q9JJY5
35 35 5 35.7 44 2 Q9PB90
36 36 5 35.7 55 1 Q9HKNO
37 37 5 35.7 78 2 O07779
38 38 5 35.7 92 9 Q9XJS6
39 39 5 35.7 99 5 Q9VX03
40 40 5 35.7 100 2 Q9KTL7
41 41 5 35.7 105 2 Q9EXM4
42 42 5 35.7 108 4 Q9PIL7
43 43 5 35.7 118 8 Q9MJ15
44 44 5 35.7 118 8 Q9MJ11
45 45 5 35.7 118 8 Q9MJ10

ALIGNMENTS

RESULT 1
ID Q9F7S4 PRELIMINARY; PRT; 139 AA.
AC Q9F7S4; 2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PREDICTED DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE.
OS uncultured proteobacterium EBAC31A08.
OC Bacteria; Proteobacteria; environmental samples.
OX NCBI_TaxID=133804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20446260; PubMed=10988064;
RA Beja O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A., Nguyen L.P.,
RA Jovanovich S.B., Gates C.M., Feldman R.A., Spudich J.L., Spudich E.N.,
RA DeLong E.F.;
RT "Bacterial rhodopsin: evidence for a new type of phototrophy in the
RT Sea";
RL Science 289:1902-1906(2000).
DR EMBL; AF279106; AAG10445.1; -.
DR InterPro; IPR001428; dUTPase.
DR Pfam; PF00692; dUTPase; 1.
DR DR ProDom; PD000946; dUTPase; 1.
KW Hydrolase.
SQ SEQUENCE 139 AA; 14883 MW; 0DCA7E61F9E16C62 CRC64;

Query Match 71.4%; Score 10; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGLIDSDYQG 10
| | | | | | | | | |
Db 73 VGLIDSDYQG 82

RESULT 2
ID Q9JZU7 PRELIMINARY; PRT; 150 AA.
AC Q9JZU7
DT 01-OCT-2000 (TReMBLrel. 15, Created)

```

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
GN NMB0893.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=MC58 / SEROGROUP B.
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Mason T.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Cotton M.D., Utterback T.R., Khuri H., Qin H., Yamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RL MC58." 287:1809-1815(2000).
RL Science; AEO02441; AAF41302.1; -.
DR TIGR; NMB0893; -.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD000946; dUTPase; 1.
DR HydroLase: Complete proteome.
KW HYDROLASE
SQ SEQUENCE 150 AA; 16129 MW; 120F426E74D37A7E CRC64;

Query Match 71.4%; Score 10; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDSDYQG 10
Db 84 VGLIDSDYQG 93

RESULT 3
ID Q9JUW1 PRELIMINARY; PRT; 150 AA.
AC Q9JUW1
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PUTATIVE DEOXYURIDINE 5'-TRIPHOSPHATE NUCLEOTIDOHYDROLASE
DE (EC 3.6.1.23)
GN DUT OR NNA112
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
RL EMBL; AL162755; CAB84374.1; -.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF00692; dUTPase; 1.
DR ProDom: PD000946; dUTPase; 1.
DR HydroLase: Complete proteome.
KW HYDROLASE
SQ SEQUENCE 150 AA; 16285 MW; 9EFB0077F7CBD12C CRC64;

Query Match 71.4%; Score 10; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDSDYQG 10
Db 85 VGLIDSDYQG 94

RESULT 5
ID Q9PGZ6 PRELIMINARY; PRT; 155 AA.
AC Q9PGZ6
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE DUTPASE.
GN XF0150.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX NCBI_TaxID=2371;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=9A5C.
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.N., Brites M.R.S.,

```

RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuranee E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,
 RT "The genome sequence of the plant pathogen *Xyella fastidiosa*."
 RL Nature 406:151-159(2000).
 DR EMBL; AE003868; AAF82963.1; -.
 DR InterPro; IPR001428; dUTPase.
 DR Pfam; PF00692; dUTPase; 1.
 DR ProDom; PD000946; dUTPase; 1.
 KW Complete proteome.
 SQ SEQUENCE 155 AA; 16248 MW; 20ADD9D48109E0CE CRC64;

Query Match: 64.3%; Score 9; DB 2; Length 155;
 Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDSXYQG 10
 |||||
 DB 90 GLIDSXYQG 98

RESULT 6
 Q9LSA1
 ID Q9LSA1 PRELIMINARY; PRT; 139 AA.
 AC Q9LSA1
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE GENOMIC DNA, CHROMOSOME 3, P1 CLONE: MVE11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
 RT clones";
 RL DNA Res. 7:131-135(2000).
 DR EMBL; AB026654; BAB01802.1; -.
 SQ SEQUENCE 139 AA; 15966 MW; 696B1DEA844218C7 CRC64;

Query Match: 42.9%; Score 6; DB 10; Length 139;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YQGQLM 13
 |||||
 DB 60 YQGQLM 65

RESULT 7
 O03945
 ID O03945 PRELIMINARY; PRT; 232 AA.
 AC O03945;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE LACTOBACILLUS BACTERIOPHAGE PHIGLE COMPLETE GENOMIC DNA.
 GN RORF232.
 OS Bacteriophage phigle.
 OC Viruses.
 OX NCBI_TaxID=52979;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97225795; PubMed=9073065;
 RA Kodaira K.I., Oki M., Kakikawa M., Watanabe N., Hirakawa M.,
 RA Yamada K., Taketo A.;
 RT "Genome structure of the Lactobacillus temperate phage phigle: the
 RT whole genome sequence and the putative promoter/repressor system.";
 RL Gene 187:45-53(1997).
 DR EMBL; X98106; CAA66755.1; -.
 SQ SEQUENCE 232 AA; 25836 MW; 6A77FBA26AFC1742 CRC64;

Query Match: 42.9%; Score 6; DB 9; Length 232;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDS 6
 |||||
 DB 155 VGLIDS 160

RESULT 8
 Q9U8P7
 ID Q9U8P7 PRELIMINARY; PRT; 279 AA.
 AC Q9U8P7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE AQPA.
 GN AQPA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA Mitra B.N., Yoshino R., Kato M., Morio T., Urushihara H., Tanaka Y.;
 RT "Sequence analysis of a water channel protein gene *aqpA* in
 RT Dictyostelium discoideum";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB032841; BAA85158.1; -.
 DR HSP; P11244; IFX8.
 DR InterPro; IPR000425; MIP.
 DR Pfam; PF00230; MIP; 1.
 DR PRINTS; PR00783; MINTRINSCP.
 DR PROSITE; PS00221; MIP; UNKNOWN1.
 SQ SEQUENCE 279 AA; 29905 MW; 1B4928E57ACC7EA7 CRC64;

Query Match: 42.9%; Score 6; DB 5; Length 279;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Query Match      42.9%; Score 6; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDS 6
Db 109 VGLIDS 114

RESULT 9
ID P95727 PRELIMINARY; PRT; 280 AA.
AC P95727
DT 01-MAY-1997 (TrEMBLrel. 03, Created).
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 1-CYCLOHEXYLCARBONYL COA REDUCTASE.
GN CHCA.
OS Streptomyces collinus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=42684;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97113449; PubMed=8955309;
RA Wang P., Denoya C.B., Morgenstern M.R., Skinner D.D., Wallace K.K.,
RA Digate R.J., Patton S., Banavali N., Schuler G., Speedie M.K.,
RA Reynolds K.A.;
RT "Cloning and characterization of the gene encoding l-
RT cyclohexenylcarbonyl coenzyme A reductase from Streptomyces
RT collinus."
RL J. Bacteriol. 178:6873-6881(1996).
CC -!- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
CC FAMILY.
DR EMBL; U72144; AAC44655.1; -.
DR HSP; P50163; 2AE1.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR002347; ADH_short_C2.
DR Pfam; PF00106; adh_short; 1.
DR Pfam; PF00678; adh_short_C2; 1.
DR Oxidoreductase.
KW OXIDOREDUCTASE.
SQ SEQUENCE 280 AA; 29849 MW; 20F8B51A21E76135 CRC64;

Query Match      42.9%; Score 6; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDS 7
Db 192 GLIDS 197

RESULT 10
ID Q9RP73 PRELIMINARY; PRT; 355 AA.
AC Q9RP73
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 41.1 KDA PROTEIN.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Mannheimia.
OX NCBI_TaxID=75985;
RN [1]
RP SEQUENCE FROM N.A.
RA Lo R.Y.C., Hills T.L., Kostrzynska M., McKerral L.J.;
RT "Analysis of the capsule biosynthetic cluster of Pasteurella
RT haemolytica A1."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RW EMBL; AF170495; AAF08246.1; -.
KW Hypothetical protein.
SQ SEQUENCE 355 AA; 41088 MW; 35DADE832699123A CRC64;

Query Match      42.9%; Score 6; DB 2; Length 388;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDS 8
Db 349 LIDS 354

RESULT 12
ID Q29700 PRELIMINARY; PRT; 404 AA.
AC Q29700
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE THREONINE SYNTHASE (THRC-1).
GN AF0551.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,

```

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 DR EMBL; AF001066; AAB90683.1; -.
 DR TIGR; AF0551; -.
 DR InterPro; IPR001926; PALP.
 DR Pfam; PF00291; PALP; 1.
 DR Hypothetical protein; Complete proteome.
 KW
 SQ SEQUENCE 404 AA; 43941 MW; 05C2E77770C26050 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 404;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGLIDS 6

Db 253 VGLIDS 258

RESULT 13

Q9NKA0 Q9NKA0 PRELIMINARY; PRT; 433 AA.
 AC Q9NKA0;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BG:DS00929.8 PROTEIN.
 GN YELLOW-C OR BG:DS00929.8 OR CG4182.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y, CN BW SP;
 RX MEDLINE=99403001; PubMed=10471707;
 RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
 RA Celnikier S., Rubin G.M.;
 RA "An exploration of the sequence of a 2.9-Mb region of the genome of
 RT Drosophila melanogaster: The Adh region.";
 RL Genetics 153:179-219(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y, CN BW SP;
 RA Celnikier S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
 RA Butenhoff C., Champagne M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
 RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
 RA Houston K.A., Humastli S.R., Karra K., Kearney L., Kim E., Lee B.,
 RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
 RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequiera A.,
 RA Sethi H., Snir E., Svirska R.R., Wan K.H., Weinburg T., Zhang R.,
 RA Zieran L.L., Rubin G.M.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003411; AAF44912.1; -.
 DR FlyBase; FBgn0041713; Yellow-c.
 SQ SEQUENCE 433 AA; 48733 MW; 93A43238B0FB7C77 CRC64;

Query Match 42.9%; Score 6; DB 5; Length 433;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLIDS 7

Db 370 GLIDS 375

RESULT 14

Q9VJQ3 Q9VJQ3 PRELIMINARY; PRT; 438 AA.
 AC Q9VJQ3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG4182 PROTEIN.
 GN YELLOW-C OR BG:DS00929.8 OR CG4182.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2183-2195(2000).
 DR EMBL; AE003646; AAF53432.1; -.
 DR FlyBase; FBgn0041713; yellow-c.
 SQ SEQUENCE 438 AA; 49245 MW; 6B3EC7553D3A58D6 CRC64;

Query Match 42.9%; Score 6; DB 5; Length 438;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLIDS 7

Db 375 GLIDS 380

RESULT 15

Q9Y7W5
 ID Q9Y7W5 PRELIMINARY; PRT; 455 AA.
 AC Q9Y7W5;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PI-SCEI PROTEIN.
 GN TFP1 OR PI-SCEI OR YDL185W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Volckaert G., Verhasselt P., Voet M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z74233; CAA98762.1; --
 DR HSP: P17255; 1VDE
 DR SGD: S0002344; TFP1
 DR InterPro: IPR003586; HintC.
 DR InterPro: IPR003587; HintN.
 DR InterPro: IPR002203; Intein.
 DR PRINTS: PR00379; INTEIN.
 DR SMART: SM00305; HintC; 1.
 DR SMART: SM00306; HintN; 1.
 DR PROSITE: PS00881; PROTEIN_SPLICING; 1.
 SO SEQUENCE 455 AA; 51034 MW; E3656F3C436EBA7 CRC64;

Query Match 42.9%; Score 6; DB 3; Length 455;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 GLIDSD 7
 Db 321 GLIDSD 326

Search completed: January 31, 2002, 13:37:49
 Job time: 174 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:16 ; Search time 78.64 seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-78
Perfect score: 14
Sequence: 1 VGLIDSDYQGQLMI 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	151	1 WPECDU	dutp pyrophosphata
2	14	100.0	151	2 D86040	deoxyuridinetripho
3	14	100.0	154	2 B84995	dutp pyrophosphata
4	10	71.4	150	2 D81877	probable dutp pyro
5	10	71.4	150	2 C81146	deoxyuridine 5'-tr
6	10	71.4	151	2 H64104	dutp pyrophosphata
7	10	71.4	151	2 C82982	deoxyuridine 5'-tr
8	10	71.4	152	2 S44300	dutp pyrophosphata
9	9	64.3	155	2 A82843	dutpase XF0150 [lm
10	6	42.9	232	2 T13226	hypothetical prote
11	6	42.9	293	2 S51571	threonine synthase
12	6	42.9	404	2 G69318	homothallic switch
13	6	42.9	586	2 S59301	homothallic switch
14	6	42.9	586	2 JC2407	nitrous-oxide redu
15	6	42.9	650	2 T44660	probable calcium c
16	6	42.9	724	2 B85045	phenylalanine--trn
17	6	42.9	805	2 S73374	lysyl aminopeptida
18	6	42.9	849	2 JU0191	H+-transporting AT
19	6	42.9	1071	1 PXBYVA	probable calcium c
20	6	42.9	1148	2 T18770	myosin X - bovine
21	6	42.9	1214	2 G70953	myosin X - human
22	6	42.9	2052	2 T18519	myosin X - mouse
23	6	42.9	2058	2 A59267	hypothetical prote
24	6	42.9	2062	2 A59297	hypothetical 6K pr
25	5	35.7	44	2 D82579	Swf9 protein - swi
26	5	35.7	53	2 JS0250	hypothetical prote
27	5	35.7	69	2 D37949	hypothetical prote
28	5	35.7	78	2 B70909	hypothetical prote
29	5	35.7	99	2 D65046	hypothetical prote

ALIGNMENTS

RESULT 1

WPECDU
dutp pyrophosphatase (EC 3.6.1.23) - Escherichia coli
C:Species: Escherichia coli
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 24-Sep-1999
C:Accession: A30388; B65165; Q00497
R:Lundberg, L.G.; Thoresson, H.O.; Karlstroem, O.H.; Nymann, P.O.
EMBO J. 2, 967-971, 1983
A:Title: Nucleotide sequence of the structural gene for dutpase of Escherichia coli K
A:Reference number: A30388; MUID:84057777
A:Accession: A30388
A:Molecule type: DNA
A:Residues: 1-151 <LON>
A:Cross-references: GB:X01714; NID:g41296; PIDN:CAA25859.1; PID:g41297
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B65165
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000441; GB:U00096; NID:g1790063; PIDN:AAC76664.1; PID:g17900
A:Experimental source: strain K-12, substrain MGL655
C:Comment: This enzyme catalyzes the hydrolysis of dutp (deoxyuridine 5'-triphosphate
ar concentration of dutp.
C:Genetics:
A:Gene: dut
A:Map position: 82 min
C:Superfamily: dutp pyrophosphatase
C:Keywords: hydrolase; nucleotide metabolism

Query Match 100.0%; Score 14; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 7.4e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGLIDSDYQGQLMI 14
Db 85 VGLIDSDYQGQLMI 98

RESULT 2

D86040
deoxyuridinetriphosphatase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: D86040
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoumis, K.; Apoda
Nature 409, 529-533, 2001

hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
probable dutp pyro
hypothetical prote
probable dutp pyro
dutp nucleotidohyd
acetyltransferase,
hypothetical prote
protein tyrosine p
spoIIIGA homolog -
hypothetical 18.7K
heat shock protein
hypothetical prote

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206951
A:Accession: D86040
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <SFO>
A:Cross-references: GB:AE005174; MID:gl2518396; PIDN:AAG58784.1; GSPDB:GN00145; UMGP:250
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Gene: dut

Query Match 100.0%; Score 14; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 7.4e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDSDYQGLMI 14
|||||
Db 85 VGLIDSDYQGLMI 98

RESULT 3
B84995
dUTP pyrophosphatase (EC 3.6.1.23) [imported] - Buchnera sp. (strain-APS)
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: B84995
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A:Reference number: A84995; A84930; MUID:20445173
A:Accession: B84995
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <SFO>
A:Cross-references: GB:AP000398; GSPDB:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: dut; BU560
C:Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 7.6e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDSDYQGLMI 14
|||||
Db 85 VGLIDSDYQGLMI 98

RESULT 4
D81877
probable dUTP pyrophosphatase (EC 3.6.1.23) NM1112 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D81877
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
Holroyd, S.; Jorgensen, K.; Leather, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556
A:Accession: D81877
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-150 <PAR>
A:Cross-references: GB:AL162755; GB:AL157959; MID:g7379742; PIDN:CAB84374.1; PID:g737980
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: dut; NM1112
C:Superfamily: retroviral proteinase
C:Keywords: hydrolase

Query Match 71.4%; Score 10; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDSDYQG 10
|||||
Db 84 VGLIDSDYQG 93

RESULT 5
C81146
deoxyuridine 5'-triphosphate nucleotidohydrolase NMB0893 [imported] - Neisseria mening
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81146
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: C81146
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-150 <JET>
A:Cross-references: GB:AE002441; GB:AE002098; MID:g7226123; PIDN:AAF41302.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0893
C:Superfamily: retroviral proteinase

Query Match 71.4%; Score 10; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDSDYQG 10
|||||
Db 84 VGLIDSDYQG 93

RESULT 6
H64104
dUTP pyrophosphatase (EC 3.6.1.23) - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 22-Jun-1999
C:Accession: H64104
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: H64104
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <TRIG>
A:Cross-references: GB:U032776; GB:L42023; MID:gl573969; PIDN:AAC22615.1; PID:gl573979
C:Superfamily: dUTP pyrophosphatase
C:Keywords: hydrolase; nucleotide metabolism

Query Match 71.4%; Score 10; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.00023;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDSDYQG 10
|||||
Db 85 VGLIDSDYQG 94

RESULT 7
 C82982
 deoxyuridine 5'-triphosphate nucleotidohydrolase PA5321 [imported] - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: C82982
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337
 A:Accession: C82982
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <STO>
 A:Cross-references: GB:AE004944; GB:AE004091; MID:9951628; PIDN:AAG08706.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: dut; PA5321

Query Match 71.4%; Score 10; DB 2; Length 151;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VGLIDSDYQG 10
 Db 85 VGLIDSDYQG 94

RESULT 8
 S44300
 dUTP pyrophosphatase (EC 3.6.1.23) - Coxiella burnetii
 C:Species: Coxiella burnetii
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Jun-1999
 C:Accession: S44300
 R:Thiele, D.; Willems, H.; Oswald, W.; Krauss, H.
 submitted to the EMBL Data Library, May 1994
 A:Reference number: S44297
 A:Accession: S44300
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-152 <THI>
 A:Cross-references: EMBL:X79075; MID:g483518; PIDN:CAA55678.1; PID:g483522
 C:Superfamily: dUTP pyrophosphatase
 C:Keywords: hydrolase; nucleotide metabolism

Query Match 71.4%; Score 10; DB 2; Length 152;
 Best Local Similarity 100.0%; Pred. No. 0.00023;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VGLIDSDYQG 10
 Db 86 VGLIDSDYQG 95

RESULT 9
 A82843
 dUTPase XF0150 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
 C:Accession: A82843
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82843
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-155 <SIM>
 A:Cross-references: GB:AE003868; GB:AE003849; MID:g9104930; PIDN:AAF82963.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Perro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr Chadô, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins chado, M.A.; Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF0150
 C:Superfamily: dUTP pyrophosphatase

Query Match 64.3%; Score 9; DB 2; Length 155;
 Best Local Similarity 100.0%; Pred. No. 0.0031;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDSDYQG 10
 Db 90 GLIDSDYQG 98

RESULT 10
 T13226
 hypothetical protein R232 - Lactobacillus phage phi-gle
 C:Species: Lactobacillus phage phi-gle
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
 C:Accession: T13226
 R:Kodaira, K.I.; Oki, M.; Kakikawa, M.; Watanabe, N.; Hirakawa, M.; Yamada, K.; Taket Gene 187, 45-53, 1997
 A:Title: Genome structure of the Lactobacillus temperate phage phi gle: the whole gen A:Reference number: Z17631; MUID:97225795
 A:Accession: T13226
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-232 <KOD>
 A:Cross-references: EMBL:X98106; MID:g1926320; PIDN:CAA66755.1; PID:g1926370
 C:Genetics:
 A:Note: Rorf232
 C:Superfamily: Lactobacillus phage phi-gle hypothetical protein R232

Query Match 42.9%; Score 6; DB 2; Length 232;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDS 6
 Db 155 VGLIDS 160

RESULT 11
 S51571
 hypothetical protein 293 - Rhizobium meliloti
 C:Species: Rhizobium meliloti
 C:Date: 15-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 08-Oct-1999
 C:Accession: S51571; S43166
 R:Rosbach, S.; Kulp, D.A.; Rosbach, U.; de Bruijn, F.J.
 Mol. Gen. Genet. 245, 11-24, 1994
 A:Title: Molecular and genetic characterization of the rhizopine catabolism (mocABRC) A:Reference number: S51569; MUID:95147842
 A:Accession: S51571
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-293 <ROS>
 A:Cross-references: EMBL:X78503; NID:g468758; PIDN:CAA55268.1; PID:g468761
 A:Experimental source: strain L5-30
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994

Query Match 42.9% Score 6; DB 2; Length 293;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDSD 7
 |||||
 Db 96 GLIDSD 101

RESULT 12
 G69318
 threonine synthase (EC 4.2.99.2) thrC-1 AF0551 [similarity] - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-May-2000
 C:Accession: G69318
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390: 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343
 A:Accession: G69318
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-404 <RA>
 A:Cross-references: GB:AE001066; GB:AE000782; NID:g2689389; PIDN:AA890683.1; PID:g265006
 C:Superfamily: threonine dehydratase
 C:Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate
 F:107/binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 42.9% Score 6; DB 2; Length 404;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLIDS 6
 |||||
 Db 253 VGLIDS 258

RESULT 13
 S59301
 homothallic switching endonuclease - yeast (Saccharomyces cerevisiae)
 N:Alternate names: HO endonuclease; homothallism protein; protein D0827; protein YDL227c
 C:Species: Saccharomyces cerevisiae
 C>Date: 08-Jul-1995 #sequence_revision 01-Dec-1995 #text_change 29-Oct-1999
 C:Accession: S59301; A25390; S67790
 R:Rah, D.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: S59301
 A:Accession: S59301
 A:Molecule type: DNA
 A:Residues: 1-586 <RA>
 A:Cross-references: EMBL:X90957; NID:g984693; PIDN:CAA62447.1; PID:g984694
 R:Russell, D.W.; Jensen, R.; Zoller, M.J.; Burke, J.; Errede, B.; Smith, M.; Herskowitz,
 Mol. Cell. Biol. 6, 4281-4294, 1986
 A:Title: Structure of the Saccharomyces cerevisiae HO gene and analysis of its upstream
 A:Reference number: A25390; MUID:87089786
 A:Accession: A25390
 A:Molecule type: DNA
 A:Residues: 1-188, 'T', 190-222, 'G', 224-404, 'L', 406-474, 'H', 476-586 <RUS>
 A:Cross-references: EMBL:M14678; NID:g171697; PIDN:AAA34683.1; PID:g171698
 R:Rasmussen, S.W.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67778

A:Accession: S67790
 A:Molecule type: DNA
 A:Residues: 1-586 <RAS>
 A:Cross-references: EMBL:Z74275; NID:gl431382; PIDN:CAA98806.1; PID:e253273; PID:gl43
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:HO
 A:Cross-references: SGD:S0002386; MIPS:YDL227c
 A:Map position: 4L
 C:Function:
 A:Description: site-specific endonuclease that cleaves a site in the MAT locus on chr
 C:Keywords: DNA binding; nucleus; zinc finger

Query Match 42.9% Score 6; DB 2; Length 586;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDSD 7
 |||||
 Db 328 GLIDSD 333

RESULT 14
 JC2407
 homothallic switching endonuclease - yeast (Saccharomyces cerevisiae) (strain wy2)
 N:Alternate names: HO endonuclease; protein YDL227c
 C:Species: Saccharomyces cerevisiae
 A:Variety: strain wy2-14-d
 C>Date: 26-May-1995 #sequence_revision 26-May-1995 #text_change 07-May-1999
 C:Accession: JC2407
 R:Tani, Y.; Kurokui, T.; Masaki, C.; Hayakawa, M.; Ekino, K.; Tomohiro, Y.; Miyata, A
 Biosci. Biotechnol. Biochem. 58, 2228-2231, 1994
 A:Title: A novel type of life cycle "delayed homothallism" in Saccharomyces cerevisia
 A:Reference number: JC2407; MUID:95128029
 A:Accession: JC2407
 A:Molecule type: DNA
 A:Residues: 1-586 <TAN>
 A:Experimental source: strain wy2-14-d
 C:Genetics:
 A:Gene: SGD:HO
 A:Cross-references: SGD:S0002386; MIPS:YDL227c
 A:Map position: 4L
 C:Function:
 A:Description: site-specific endonuclease that cleaves a site in the MAT locus on chr
 C:Keywords: DNA binding; nucleus; zinc finger
 F:466-489/Region: zinc finger CCCC motif
 F:558-578/Region: zinc finger CCHC motif

Query Match 42.9% Score 6; DB 2; Length 586;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDSD 7
 |||||
 Db 328 GLIDSD 333

RESULT 15
 T44660
 nitrous-oxide reductase (EC 1.7.99.6) [imported] - Bradyrhizobium japonicum
 C:Species: Bradyrhizobium japonicum
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
 C:Accession: T44660
 R:Bedmar, E.J.; Velasco, L.; Xu, C.A.; Delgado, M.J.
 submitted to the EMBL Data Library, March 1998
 A:Description: Bradyrhizobium japonicum nosRZDFLX gene cluster.
 A:Reference number: Z22825
 A:Accession: T44660
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-650 <BED>

A:Cross-references: EMBL:AJ002531; PIDN:CAA05518.1
A:Experimental source: strain USDA110
C:Genetics:
A:Gene: nosZ
C:Superfamily: nitrous-oxide reductase
C:Keywords: oxidoreductase

Query Match 42.9%; Score 6; DB 2; Length 650;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DSDYQG 10
|||||
Db 260 DSDYQG 265

Search completed: January 31, 2002, 13:20:17
Job time: 118 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:14 ; Search time 65.13 Seconds
(without alignments)
4.837 Million cell updates/sec

Title: US-08-957-709-78

Perfect score: 14

Sequence: 1 VGLIDSDYQGQLMI 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	151	2	US-08-824-405-7
2	6	42.9	340	4	US-08-975-762-54
3	6	42.9	340	4	US-09-295-028-54
4	6	42.9	340	4	US-09-106-582-54
5	5	35.7	27	2	US-08-563-892A-16
6	5	35.7	50	1	US-07-903-456-5
7	5	35.7	67	4	US-09-025-151-27
8	5	35.7	142	1	US-08-307-499-22
9	5	35.7	142	4	US-09-299-268-22
10	5	35.7	153	2	US-08-387-942C-43
11	5	35.7	175	4	US-09-357-251-2
12	5	35.7	212	1	US-08-462-965A-2
13	5	35.7	212	1	US-08-462-169B-21
14	5	35.7	212	3	US-09-103-079-21
15	5	35.7	216	2	US-08-821-637-3
16	5	35.7	261	2	US-08-477-451-40
17	5	35.7	297	2	US-08-481-956A-10
18	5	35.7	297	2	US-08-629-291A-10
19	5	35.7	297	2	US-08-658-335B-10
20	5	35.7	383	4	US-08-857-076-105
21	5	35.7	497	1	US-08-278-635B-5
22	5	35.7	497	3	US-08-464-258B-5
23	5	35.7	497	3	US-08-471-961-5
24	5	35.7	548	2	US-08-871-266B-17
25	5	35.7	548	2	US-09-018-864A-17
26	5	35.7	548	3	US-08-871-267B-23
27	5	35.7	548	4	US-09-618-419-23

28	5	35.7	622	4	US-09-305-381-2	Sequence 2, Appli
29	5	35.7	623	1	US-08-734-925-2	Sequence 2, Appli
30	5	35.7	630	3	US-08-771-986A-2	Sequence 2, Appli
31	5	35.7	630	3	US-08-769-802A-2	Sequence 2, Appli
32	5	35.7	666	4	US-08-982-785A-11	Sequence 11, Appli
33	5	35.7	776	1	US-08-021-601-2	Sequence 2, Appli
34	5	35.7	776	1	US-08-082-849B-2	Sequence 2, Appli
35	5	35.7	776	5	PCT-US94-01624-2	Sequence 2, Appli
36	5	35.7	823	1	US-08-461-551-2	Sequence 2, Appli
37	5	35.7	849	1	US-08-405-392-18	Sequence 18, Appli
38	5	35.7	849	3	US-08-487-691-18	Sequence 18, Appli
39	5	35.7	866	1	US-08-405-392-17	Sequence 17, Appli
40	5	35.7	866	3	US-08-487-691-17	Sequence 17, Appli
41	5	35.7	884	1	US-07-718-575-12	Sequence 12, Appli
42	5	35.7	884	1	US-08-481-206-12	Sequence 12, Appli
43	5	35.7	884	2	US-08-486-269A-12	Sequence 12, Appli
44	5	35.7	905	1	US-08-405-392-2	Sequence 2, Appli
45	5	35.7	905	3	US-08-487-691-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-824-405-7
; Sequence 7, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladhner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-824-405-7

Query Match 100.0%; Score 14; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 4.2e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGLIDSDYQGQGLMI 14
| | | | | | | | | | | | | | | |
Db 85 VGLIDSDYQGQGLMI 98

RESULT 2
US-09-975-762-54
; Sequence 54, Application US/08975762
; Patent No. 6207169
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/975,762
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-975-762-54

Query Match 42.9%; Score 6; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 SDYQGG 11
| | | | | | | | | | | | | | | |
Db 2 SDYQGG 7

RESULT 3
US-09-295-028-54
; Sequence 54, Application US/09295028
; Patent No. 6277381
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.439C4
; CURRENT APPLICATION NUMBER: US/09/295,028
; CURRENT FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 85

; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 54
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Ehrlichia sp.
US-09-295-028-54

Query Match 42.9%; Score 6; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 SDYQGG 11
| | | | | | | | | | | | | | | |
Db 2 SDYQGG 7

RESULT 4
US-09-106-582-54
; Sequence 54, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/106,582
; FILING DATE: 29-JUN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-106-582-54

Query Match 42.9%; Score 6; DB 4; Length 340;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 SDYQGG 11
| | | | | | | | | | | | | | | |
Db 2 SDYQGG 7

RESULT 5
US-08-563-892A-16
; Sequence 16, Application US/08563892A
; Patent No. 5976819

THERAP

GENERAL INFORMATION:
APPLICANT: Finkel, Terri H.
APPLICANT: Rozdzial, Moshe M.
TITLE OF INVENTION: PRODUCT AND PROCESS TO REGULATE ACTIN
POLYMERIZATION IN T LYMPHOCYTES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/563,892A
FILING DATE: 21-NOV-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-33
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-563-892A-16

Query Match 35.7%; Score 5; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DSDYQ 9
Db 10 DSDYQ 14

RESULT 6
US-07-903-456-5
Sequence 5, Application US/07903456
Patent No. 5574144
GENERAL INFORMATION:
APPLICANT: KAMBOJ, Rajender
APPLICANT: ELLIOTT, Candace
APPLICANT: NUTT, Stephen
TITLE OF INVENTION: KAINATE-BINDING HUMAN CNS RECEPTORS OF
THE EAA4 FAMILY
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/903,456

FILING DATE: 19920624
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16777/183/ALLE
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
US-07-903-456-5

Query Match 35.7%; Score 5; DB 1; Length 50;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDS 6
Db 27 GLIDS 31

RESULT 7
US-09-025-151-27
Sequence 27, Application US/09025151
Patent No. 6187535
GENERAL INFORMATION:
APPLICANT: Legrain, Pierre
APPLICANT: Fromont, Micheline
APPLICANT: Rain, Jean-Christophe
TITLE OF INVENTION: FAST AND EXHAUSTIVE METHOD FOR SELECTING A PREY
POLYPEPTIDE INTERACTING WITH A BAIT POLYPEPTIDE OF
INTEREST: APPLICATION TO THE CONSTRUCTION OF MAPS OF
INTERACTORS POLYPEPTIDES
FILE REFERENCE: 03495-0164
CURRENT APPLICATION NUMBER: US/09/025,151
CURRENT FILING DATE: 1998-02-18
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 27
LENGTH: 67
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-025-151-27

Query Match 35.7%; Score 5; DB 4; Length 67;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGLID 5
Db 60 VGLID 64

RESULT 8
US-08-307-499-22
Sequence 22, Application US/08307499
Patent No. 5651972
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Vi uela, Eladio
APPLICANT: Gibbs, E.P.J.
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
Live Vaccine Vector
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
US-08-307-499-22

```

; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,499
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 142 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-307-499-22

Query Match 35.7%; Score 5; DB 1; Length 142;
Best Local Similarity 100.0%; Pred.No.1.le+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 IDSDY 8
Db 80 IDSDY 84

RESULT 9
US-09-299-268-22
; Sequence 22, Application US/09299268
; Patent No. 6217882
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vialla, Etadio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; TITLE OF INVENTION: Live Vaccine Vector
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,268
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/901,127
; FILING DATE:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 142 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-268-22

Query Match 35.7%; Score 5; DB 4; Length 142;
Best Local Similarity 100.0%; Pred.No.1.le+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 IDSDY 8
Db 80 IDSDY 84

RESULT 10
US-08-387-942C-43
; Sequence 43, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELGA
; APPLICANT: VALLA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O.BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,942C
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M.
; REGISTRATION NUMBER: 28,977

```


; REFERENCE/DOCKET NUMBER: 1809-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-387-942C-43

Query Match 35.7%; Score 5; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YQQL 12
Db 138 YQQL 142
|||||

RESULT 11
US-09-357-251-2
; Sequence 2, Application US/09357251
; Patent No. 6271441
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Schwaber, James S.
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase
; FILE REFERENCE: BB-1193
; CURRENT APPLICATION NUMBER: US/09/357,251
; CURRENT FILING DATE: 1999-07-20
; EARLIER APPLICATION NUMBER: 60/093,530
; EARLIER FILING DATE: July 21, 1998
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 175
; TYPE: PRT
; ORGANISM: zea mays
US-09-357-251-2

Query Match 35.7%; Score 5; DB 4; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLIDS 6
Db 21 GLIDS 25
|||||

RESULT 12
US-08-462-965A-2
; Sequence 2, Application US/08462965A
; Patent No. 5728546
; GENERAL INFORMATION:
; APPLICANT: Greene Ph.D., John M.
; APPLICANT: Gruber Ph.D., Joachim R.
; APPLICANT: Rosen, Craig R.
; TITLE OF INVENTION: Fibroblast Growth Factor 13
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: US

; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,965A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-439 (PF171)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 212 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-462-965A-2

Query Match 35.7%; Score 5; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YQQL 12
Db 175 YQQL 179
|||||

RESULT 13
US-08-462-169B-21
; Sequence 21, Application US/08462169B
; Patent No. 5773252
; GENERAL INFORMATION:
; APPLICANT: John Greene and Craig A. Rosen
; TITLE OF INVENTION: Fibroblast Growth Factor-15
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,169B
; FILING DATE: 05 JUN 95
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-441 (PF203)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 212 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR

```
; MOLECULE TYPE: PROTEIN
US-08-462-169B-21

Query Match      35.7%; Score 5; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 YQGQL 12
      |||||
Db      175 YQGQL 179

RESULT 14
US-09-103-079-21
; Sequence 21, Application US/09103079A
; Patent No. 6013477
; GENERAL INFORMATION:
; APPLICANT: Greene, John M.
; APPLICANT: Rosen, Craig A.
; TITLE OF INVENTION: Fibroblast Growth Factor 15
; FILE REFERENCE: PF203D1
; CURRENT APPLICATION NUMBER: US/09/103,079A
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 08/462,169
; EARLIER FILING DATE: 1995-06-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-103-079-21

Query Match      35.7%; Score 5; DB 3; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 YQGQL 12
      |||||
Db      175 YQGQL 179

RESULT 15
US-08-821-637-3
; Sequence 3, Application US/08821637
; Patent No. 5912327
; GENERAL INFORMATION:
; APPLICANT: LI, YULING
; APPLICANT: OELKUNT, MARK
; APPLICANT: GENTZ, REINER L.
; TITLE OF INVENTION: METHOD OF PURIFYING CHEMOKINES FROM
; TITLE OF INVENTION: INCLUSION BODIES (AS AMENDED)
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,637
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:12 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-78

Perfect score: 14
Sequence: 1 VGLIDSDYQGQLM1 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	14	100.0	14	19	AAW72854
2	6	42.9	32	20	AAV60355
3	6	42.9	66	22	AA16091
4	6	42.9	66	22	AA28582
5	6	42.9	66	22	AAW03819
6	6	42.9	81	18	AAW28248
7	6	42.9	104	19	AAW38508
8	6	42.9	106	22	AAW20396
9	6	42.9	106	22	AAW34854
10	6	42.9	107	22	AAW15834
11	6	42.9	107	22	AAW28344

12	6	42.9	107	22	AAW03572	Peptide #2254 enco
13	6	42.9	132	22	AAW91790	C glutamicum prote
14	6	42.9	133	21	AAW56812	Arabidopsis thalia
15	6	42.9	133	21	AAW59567	Arabidopsis thalia
16	6	42.9	139	21	AAW56811	Arabidopsis thalia
17	6	42.9	139	21	AAW59566	Arabidopsis thalia
18	6	42.9	280	21	AAW23251	Streptomyces colli
19	6	42.9	340	19	AAW82639	Ehrlichia sp. exte
20	6	42.9	340	21	AAV78561	Ehrlichia antigen
21	6	42.9	388	21	AAV70155	Staphylococcus aur
22	6	42.9	586	16	AAW72471	Saccharomyces uvar
23	6	42.9	586	16	AAW72470	Saccharomyces carl
24	6	42.9	763	22	AAW91337	C glutamicum prote
25	6	42.9	848	17	AAW02205	Lactococcus lactis
26	6	42.9	869	21	AAW42020	Human ORFX ORF1784
27	6	42.9	881	20	AAV11681	Sulfated fucose-co
28	5	35.7	6	21	AAW24303	Prostate tumour as
29	5	35.7	21	16	AAW80001	ARH1 peptide mimet
30	5	35.7	26	14	AAW33844	Valpha12.1 / Jalph
31	5	35.7	26	19	AAW56144	T cell receptor al
32	5	35.7	27	21	AAW52329	Multisubunit immun
33	5	35.7	37	22	AAW31362	Peptide #5399 enco
34	5	35.7	46	21	AAW38061	Fragment of human
35	5	35.7	50	15	AAW45368	Excitatory amino a
36	5	35.7	67	20	AAW33818	YJR022W a yeast pe
37	5	35.7	93	22	AAW18062	Peptide #4496 enco
38	5	35.7	93	22	AAW30574	Peptide #4611 enco
39	5	35.7	97	22	AAW76982	Human colon cancer
40	5	35.7	101	20	AAW34660	Chlamydia pneumoni
41	5	35.7	102	19	AAW98755	H. pylori GHPO 107
42	5	35.7	116	21	AAW25167	Eucalyptus grandis
43	5	35.7	124	22	AAW32632	Peptide #6669 enco
44	5	35.7	136	22	AAW82817	S. epidermidis ope
45	5	35.7	142	18	AAW26421	Swinepox virus Hin

ALIGNMENTS

RESULT 1
AAW72854
ID AAW72854 standard; Peptide; 14 AA.
XX
AC AAW72854;
XX
DT 01-MAR-1999 (first entry)
XX
DE Escherichia coli dUTPase uridine-binding motif.
DE Polymerase enhancing factor; PEF; dCTP deaminase; PCR;
KW amplification; sequencing; replication.
XX
OS Escherichia coli.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX
(STRA-) STRATAGENE.
XX
PA Hansen CJ, Hogrefe H;
XX
PI WPI; 1998-542284/46.
XX
PT Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication
XX

PS Claim 71; Page 47; 16lpp; English.

XX This is the uridine-binding motif of the dUTPase of Escherichia
 CC coli. Sequences are provided (see AAW72849-57) of the uridine-binding
 CC motifs of dUTPases and dUTP deaminases of Pyrococcus furiosus (see
 CC AAW72847), Methanococcus jannaschii, Desulfurolobus ambivalens,
 CC Escherichia coli, yeast, human and herpesvirus; a consensus (see
 CC AAW72848) is also provided. A claimed method of enhancing a nucleic
 CC acid polymerase reaction comprises performing the reaction in the
 CC presence of one or more of the following: a polymerase enhancing
 CC factor (PEF), a dUTPase, a protein that turns-over dUTP and a
 CC protein having one or more of the sequences provided in AAW72848-57.
 CC A claimed protein having PEF activity comprises one or more of
 CC sequences given in AAW72848-57. Kits are provided for replicating
 CC nucleic acids, for site-directed mutagenesis, for nucleic acid
 CC sequencing or for amplification (preferably PCR or RT-PCR).

XX Sequence 14 AA;

Query Match 100.0%; Score 14; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.4e-08; Mismatches 0; Gaps 0;
 Matches 14; Conservative 0; Indels 0;

QY 1 VCLIDSDYOGQLMI 14
 DB 1 VGLIDSDYGGQLMI 14

RESULT 2

AAI60355
 ID AAI60355 standard; Protein: 32 AA.

XX AC AAI60355;

XX DT 31-JAN-2000 (first entry)

XX DE Human normal bladder tissue EST encoded protein 27.

XX KW Human; bladder; treatment; EST; expressed sequence tag; cytostatic;
 XX cancer; gene therapy.

XX OS Homo sapiens.

XX FN DE19818620-A1.

XX PD 28-OCT-1999.

XX PF 21-APR-1998; 98DE-1018620.

XX PR 21-APR-1998; 98DE-1018620.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

XX DR WPI; 1999-602416/52.

XX DR N-PSDB; AAZ42153.

XX PT New polypeptides and their nucleic acids, useful for treatment of
 XX bladder tumour and identification of therapeutic agents -

XX PS Claim 23; Page 258; 366pp; German.

XX This invention describes novel polypeptide fragment sequences (I) and
 CC their encoding nucleic acids (II) which are highly expressed in normal
 CC bladder tissue and have cytostatic activity. (II) are used for
 CC recombinant expression of (I) and to isolate complete genes. (I) are
 CC used to identify agents suitable for the treatment of bladder tumours,
 CC to directly treat this form of cancer (including expression from gene
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)
 CC is also used for the generation of specific antibodies. (II) are
 CC identified by assembling ESTs (expressed sequence tags) from a

CC particular tissue type before comparison of expression patterns. This
 CC allows a significantly longer fragment of the gene to be revealed, and
 CC therefore reduces the number of failures because of ESTs from different
 CC libraries representing different parts of the same unknown gene
 CC distorting the estimated frequency of occurrence in a particular tissue.
 CC AAY60329-Y60591 represent protein fragments encoded by the human normal
 CC bladder tissue cDNA library derived EST fragments represented in
 CC AAZ42122-Z42248.

XX Sequence 32 AA;

Query Match 42.9%; Score 6; DB 20; Length 32;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LIDSDY 8
 DB 24 LIDSDY 29

RESULT 3

AAI6091
 ID AAI6091 standard; Protein: 66 AA.

XX AC AAI6091;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #2525 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
 XX cervical cancer.

XX OS Homo sapiens.

XX FN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human cervical epithelial cells -

XX PS Claim 27; SEQ ID NO 20917; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs; see AAI10068-AA128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note the sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 66 AA;

Query Match 42.9%; Score 6; DB 22; Length 66;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDSQYQ 9
 |||||
 Db 51 idsdyq 56

RESULT 4

AAW28582
 ID AAW28582 standard; Protein; 66 AA.

XX AC AAW28582;

XX DT 17-OCT-2001 (first entry)

XX DE Peptide #2619 encoded by probe for measuring placental gene expression.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;
 XX KW genetic disorder.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human placenta -
 XX PS Claim 27; SEQ ID No 28851; 654pp; English.

XX SQ Sequence 66 AA;

Query Match 42.9%; Score 6; DB 22; Length 66;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDSQYQ 9
 |||||
 Db 51 idsdyq 56

RESULT 5

AAW03819

ID AAW03819 standard; Protein; 66 AA.

XX AC AAW03819;

XX DT 09-OCT-2001 (first entry)

XX DE Peptide #2501 encoded by probe for measuring breast gene expression.

XX KW Probe; human; breast disease; breast cancer; development disorder;
 XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX OS Homo sapiens.

XX PN WO200157270-A2.

XX PD 09-AUG-2001.

XX PF 29-JAN-2001; 2001WO-US00661.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-476286/51.

XX PT Novel single exon nucleic acid probe used to measuring gene expression
 XX PT in a human breast -

XX PS Claim 27; SEQ ID No 12559; 322pp; English.

XX CC The present invention relates to novel single exon nucleic acid probes
 XX CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one
 XX CC such probe. The probes are useful for measuring human gene expression in
 XX CC a human breast sample, where the probe hybridises at high stringency to a
 XX CC nucleic acid expressed in the human breast. The probes are useful for
 XX CC predicting, diagnosing, grading, staging, monitoring and prognosing
 XX CC diseases of the human breast, particularly those diseases with polygenic
 XX CC aetiology. The diseases include: breast cancer, disorders of development,
 XX CC inflammatory diseases of the breast, fibrocystic changes, proliferative,
 XX CC breast disease and non-carcinoma tumours.

XX CC Note: The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 66 AA;

Query Match 42.9%; Score 6; DB 22; Length 66;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDSQYQ 9
 |||||
 Db 51 idsdyq 56

RESULT 6

AAW28248

ID AAW28248 standard; Protein; 81 AA.

XX AC AAW28248;

XX DT 07-SEP-1998 (first entry)

XX DE Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome.
 XX
 OS Staphylococcus aureus.
 XX
 XX WO9730070-A1.
 PN 21-AUG-1997.
 XX
 XX 19-FEB-1997; 97WO-US02318.
 XX
 XX 20-FEB-1996; 96US-0011888.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 XX WPI: 1997-424969/39.
 DR N-PSDB; AAT84155.
 DR
 XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection
 XX
 XX Claim 6; Page 556; 989pp; English.
 PS
 XX The present sequence represents a Staphylococcus aureus protein of
 CC unknown function. The DNA sequence was isolated from a library of
 CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
 CC be used in the construction of ribozymes and antisense sequences to
 CC control the expression of Staphylococcal genes. The DNA sequence is
 CC also useful as a source of regulatory elements for the control of
 CC bacterial gene expression. The present protein may be used to produce
 CC vaccines to enable a host to produce specific antibodies with
 CC antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 XX
 SQ Sequence 81 AA;

Query Match 42.9%; Score 6; DB 18; Length 81;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LIDSY 8
 Db 42 lidsy 47
 |||||
 |||||

RESULT 7
 AAW38508
 ID AAW38508 standard; Protein; 104 AA.
 XX
 AC AAW38508;

DT 06-NOV-1998 (first entry)

DE S. pneumoniae ISL2 protein.

XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis.
 XX
 OS Streptococcus pneumoniae.

PN WO9743303-A1.
 XX 20-NOV-1997.
 XX
 XX 14-MAY-1997; 97WO-US07950.
 XX
 XX 14-MAY-1996; 96US-0017670.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 XX Stodola RK;
 XX WPI: 1998-008793/01.
 XX N-PSDB; AAT98575.
 DR
 XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 XX
 XX Claim 12; Pages 292-293; 483pp; English.
 PS
 XX This sequence represents a Streptococcus pneumoniae protein that, based
 CC on homology with a lactobacillus helveticus protein, is an ISL2 protein,
 CC and is encoded by a DNA sequence of the invention. The DNA sequences
 CC were isolated from S. pneumoniae strain 0100993 (NCIMB 40794). The
 CC Streptococcus pneumoniae proteins of the invention can be used to
 CC identify compounds which interact with and inhibit or activate the
 CC activity of the proteins. Antagonists can be used to treat diseases
 CC caused by S. pneumoniae proteins, through genetic immunisation. They can
 CC also be used to induce an immunological response in a mammal by
 CC inoculation with the S. pneumoniae proteins or delivery of the encoding
 CC nucleic acids in a vector adequate to produce antibody and/or T cell
 CC immune responses to protect the animal from disease. The proteins can
 CC also be used to identify antimicrobial compounds which are capable of
 CC inhibiting their bioactivity. In particular the proteins of the invention
 CC can be used to prevent adhesion of bacteria to mammalian extracellular
 CC matrix proteins on in-dwelling devices or in wounds, to block protein-
 CC mediated mammalian cell invasion, and to block the normal progression of
 CC pathogenesis in infections initiated other than by the implantation of
 CC in-dwelling devices or other surgical techniques.
 XX
 SQ Sequence 104 AA;

Query Match 42.9%; Score 6; DB 19; Length 104;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DSDYQG 10
 Db 20 dsdygg 25
 |||||
 |||||

RESULT 8
 AAM20396
 ID AAM20396 standard; Protein; 106 AA.
 XX
 AC AAM20396;

DT 12-OCT-2001 (first entry)

DE Peptide #6830 encoded by probe for measuring cervical gene expression.
 XX Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 KW
 XX Homo sapiens.
 OS
 XX WO200157278-A2.
 PN
 XX 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000670.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488901/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human cervical epithelial cells -
 XX PS Claim 27; SEQ ID No 25222; 487pp; English.
 XX CC The present invention relates to human single exon nucleic acid probes
 XX CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
 XX CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
 XX CC can be used to produce a single exon microarray, which can be used for
 XX CC measuring human gene expression in a sample derived from human cervical
 XX CC epithelial cells. By measuring gene expression, the probes are therefore
 XX CC useful in grading and/or staging of diseases of the cervix, notably
 XX CC cervical cancer.
 XX CC Note: The sequence data for this patent did not form part of the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 106 AA;

Query Match 42.9%; Score 6; DB 22; Length 106;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDS DYQ 9
 Db |||||
 92 idsdyq 97

RESULT 9
 AAM34854
 ID AAM34854 standard; Protein; 106 AA.
 XX AC AAM34854;
 XX DT 17-OCT-2001 (first entry)
 XX DE Peptide #8891 encoded by probe for measuring placental gene expression.
 XX DE Probe: microarray; human; placenta; antenatal diagnosis;
 XX DE genetic disorder.
 XX KW Homo sapiens.
 XX OS WO200157272-A2.
 XX PN 09-AUG-2001.
 XX PD 30-JAN-2001; 2001WO-US000663.
 XX PF 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488997/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human placenta -
 XX PS Claim 27; SEQ ID No 35123; 654pp; English.
 XX CC The present invention relates to single exon nucleic acid probes (SENP:
 XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
 XX CC such probe. The probes are useful for producing a microarray for
 XX CC predicting, measuring and displaying gene expression in samples derived
 XX CC from human placenta. The probes are useful for antenatal diagnosis of
 XX CC human genetic disorders.
 XX SQ Sequence 106 AA;

Query Match 42.9%; Score 6; DB 22; Length 106;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDS DYQ 9
 Db |||||
 92 idsdyq 97

RESULT 10
 AAM15834
 ID AAM15834 standard; Protein; 107 AA.
 XX AC AAM15834;
 XX DT 12-OCT-2001 (first entry)
 XX DE Peptide #2268 encoded by probe for measuring cervical gene expression.
 XX DE Probe: human; microarray; gene expression; cervical epithelial cell;
 XX DE cervical cancer.
 XX KW Homo sapiens.
 XX OS WO200157278-A2.
 XX PN 09-AUG-2001.
 XX PD 30-JAN-2001; 2001WO-US000670.
 XX PF 04-FEB-2000; 2000US-0180312.
 XX PR 26-MAY-2000; 2000US-0207456.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-0234687.
 XX PR 27-SEP-2000; 2000US-0236359.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488901/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 XX PT analyzing gene expression in human cervical epithelial cells -
 XX PS Claim 27; SEQ ID No 20660; 487pp; English.
 XX CC The present invention relates to human single exon nucleic acid probes

CC (SNP: see AA110068-AA128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 107 AA;

Query Match 42.9%; Score 6; DB 22; Length 107;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDS DYQ 9
 Db 92 idsdyq 97

RESULT 11
 AAM28344
 ID AAM28344 standard; Protein; 107 AA.

XX AC AAM28344;

XX DT 17-OCT-2001 (first entry)

XX Peptide #2381 encoded by probe for measuring placental gene expression.
 DE
 XX Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 KW
 XX Homo sapiens.
 OS

XX WO200157272-A2.
 PN
 XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-438897/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 XX Claim 27; SEQ ID No 28613; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:
 CC see AA131315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.

XX Sequence 107 AA;

Query Match 42.9%; Score 6; DB 22; Length 107;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDS DYQ 9
 Db 92 idsdyq 97

RESULT 12
 AAM03572
 ID AAM03572 standard; Protein; 107 AA.

XX AC AAM03572;

XX DT 09-OCT-2001 (first entry)

XX Peptide #2254 encoded by probe for measuring breast gene expression.

XX Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression
 in a human breast -
 XX
 XX Claim 27; SEQ ID No 12312; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes
 CC (see AA100010-AA110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 107 AA;

Query Match 42.9%; Score 6; DB 22; Length 107;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDS DYQ 9
 Db 92 idsdyq 97

RESULT 13

AAG91790
ID AAG91790 standard; Protein; 132 AA.

XX AC AAG91790;
XX DT 26-SEP-2001 (first entry)
XX DE C glutamicum protein fragment SEQ ID NO: 5544.
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis.
XX OS Corynebacterium glutamicum.

XX PN EP1108790-A2.
XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.
XX PR 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX DR WPI: 2001-376931/40.
XX DR N-PSDB; AAH67009.

XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PS Claim 17; SEQ ID NO: 5544; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a
XX CC mutant of coryneform bacterium, measuring expression amount and
XX CC analysing the expression profile or expression pattern of a gene derived
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived
XX CC from coryneform bacterium. Coryneform bacteria are useful for producing
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX CC particularly L-lysine. The present sequence is a protein described
XX CC in the exemplification of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC European Patent Office.

XX SQ Sequence 132 AA;

Query Match 42.9%; Score 6; DB 22; Length 132;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GLIUSD 7
Db 36 glidsd 41

RESULT 14

AAG56812
ID AAG56812 standard; Protein; 133 AA.

XX AC AAG56812;

XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 73116.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.

KW termination sequence.
OS Arabidopsis thaliana.
XX EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 03-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.9%; Score 6; DB 21; Length 133;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 YOGOLM 13
Db 54 Yggqlm 59

Search completed: January 31, 2002, 13:18:13
Job time: 174 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:49 ; Search time 130.99 Seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-79
Perfect score: 14
Sequence: 1 AGVVDRTYGEVKV 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.invertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	50.0	148	Q9RPA4	Q9RPA4 streptomyce
2	7	50.0	361	Q19635	Q19635 caenorhabdi
3	7	50.0	749	Q04387	Q04387 chlamydomon
4	7	50.0	1224	Q921F8	Q921F8 staphylococ
5	6	42.9	107	Q30293	Q30293 archaeglob
6	6	42.9	113	Q42461	Q42461 squalus aca
7	6	42.9	113	Q91BH9	Q91BH9 squalus aca
8	6	42.9	132	Q70107	Q70107 bacillus ce
9	6	42.9	132	Q86214	Q86214 exigubacte
10	6	42.9	132	Q9F4C7	Q9F4C7 bacillus li
11	6	42.9	132	Q9F4C0	Q9F4C0 bacillus ma
12	6	42.9	206	Q54602	Q54602 halobacteri
13	6	42.9	279	Q9EVM9	Q9EVM9 pseudomonas
14	6	42.9	287	Q30655	Q30655 azotobacter
15	6	42.9	317	Q9CKL0	Q9CKL0 pasteurella
16	6	42.9	355	Q59731	Q59731 schizosacch
17	6	42.9	376	Q9KUHO	Q9KUHO vibrio chol
18	6	42.9	379	Q9I4F1	Q9I4F1 pseudomonas
19	6	42.9	384	Q9VE94	Q9VE94 drosophila

20	6	42.9	386	2	Q9FBQ3	Q9fbq3 streptomyce
21	6	42.9	387	3	Q9P387	Q9p387 neurospora
22	6	42.9	394	2	Q9HVF2	Q9hyf2 pseudomonas
23	6	42.9	404	10	Q9AWJ4	Q9awj4 oryza sativ
24	6	42.9	406	10	Q9SN08	Q9sn08 arabidopsis
25	6	42.9	421	2	Q99SF8	Q99sf8 staphylococ
26	6	42.9	438	2	Q9L6G4	Q9l6g4 pseudomonas
27	6	42.9	619	1	Q93738	Q93738 pyrococcus
28	6	42.9	621	1	Q9UZE9	Q9uze9 pyrococcus
29	6	42.9	625	1	Q74007	Q74007 pyrococcus
30	6	42.9	802	2	Q68595	Q68595 pseudomonas
31	6	42.9	802	2	Q91648	Q91648 pseudomonas
32	6	42.9	867	2	Q9KIB1	Q9kib1 porphyromon
33	6	42.9	949	2	Q9KU47	Q9ku47 vibrio chol
34	6	42.9	1028	2	Q9A4W6	Q9a4w6 caulobacter
35	6	42.9	1091	13	Q9YH41	Q9yh41 gallus gall
36	6	42.9	1132	4	Q78418	Q78418 homo sapien
37	6	42.9	1462	4	Q9GZY2	Q9gzy2 homo sapien
38	6	42.9	1588	4	Q9GZU2	Q9gzuz2 homo sapien
39	6	42.9	1802	6	Q28633	Q28633 oryctolagus
40	6	42.9	1983	2	Q9CJ55	Q9cj55 lactococcus
41	6	42.9	2000	6	Q97791	Q97791 oryctolagus
42	6	42.9	2164	13	Q9IAR9	Q9iar9 gallus gall
43	6	42.9	3071	10	Q9SND0	Q9snd0 arabidopsis
44	6	42.9	4162	13	Q98918	Q98918 gallus gall
45	6	42.9	26926	4	Q10466	Q10466 homo sapien

ALIGNMENTS

RESULT 1

Q9RPA4: ID Q9RPA4 PRELIMINARY; PRT; 148 AA.

AC Q9RPA4; DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOHETICAL 16.3 KDA PROTEIN.

OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=T#2717;

RA Treifer A., Hoffmeister D., Westrich L., Stockert S., Weitnauer L.,
RA Fuchser J., Bindseil K., Kuenzel J., Bechthold A.;

RT "Function of glycosyl transferase genes involved in the biosynthesis
RT of urdamycin A";

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=T#2717;

RA Faust B., Westrich L., Bechthold A.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF164961; AAF00212.1; -

DR InterPro; IPR002577; DUF24.

DR Pfam; PF01638; DUF24; 1.

DR ProDom; PD004032; DUF24; 1.

KW Hypothetical protein.

SQ SEQUENCE 148 AA; 16268 MW; 21A3B2A2E55168B2 CRC64;

Query Match 50.0%; Score 7; DB 2; Length 148;

Best Local Similarity 100.0%; Pred. No. 8.7;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGVVDRTD 7

Db 80 AGVVDRTD 86

```

RESULT 2
Q19635 ID Q19635 PRELIMINARY; PRT; 361 AA.
AC Q19635;1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE F20CS.4 PROTEIN.
GN F20CS.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RA Matthews P.;
RP SEQUENCE FROM N.A.
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighthning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Shalson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
DR EMBL: 268161; CAA92297.1;
SQ SEQUENCE 361 AA; 41249 MW; F83965FC6C1B42C6 CRC64;

Query Match 50.0%; Score 7; DB 5; Length 361;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYTGEVK 13
DB 107 DYTGEVK 113

RESULT 3
O04387 ID O04387 PRELIMINARY; PRT; 749 AA.
AC O04387;1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MR 105,000 OUTER DYNEIN ARM DOCKING COMPLEX PROTEIN.
GN ODA3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97311076; PubMed=9166407;
RA Koutoulis A., Pazour G.J., Wilkerson C.G., Inaba K., Sheng H.,
RA Takada S., Witman G.B.;
RT "The Chlamydomonas reinhardtii ODA3 gene encodes a protein of the
RL outer dynein arm docking complex.";
RL J. Cell Biol. 137:1069-1080(1997).
DR EMBL: AF001309; AAC49732.1;
SQ SEQUENCE 749 AA; 83376 MW; 8ED44083EB6DF6B CRC64;

Query Match 50.0%; Score 7; DB 10; Length 749;
Best Local Similarity 100.0%; Pred. No. 39;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYTGEVK 13
DB 107 DYTGEVK 113

RESULT 5
O30293 ID O30293 PRELIMINARY; PRT; 107 AA.
AC O30293;1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF2377.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterdon S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;

```

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVDRDY 8
DB 722 GVDRDY 728

RESULT 4
Q92IF8 ID Q92IF8 PRELIMINARY; PRT; 1224 AA.
AC Q92IF8;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NARG PROTEIN.
GN NARG.
OS Staphylococcus carnosus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1281;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=TM300;
RA Pantel I., Lindgren P.E., Neubauer H., Goetz F.;
RT "Identification and characterization of the Staphylococcus carnosus
RL nitrate reductase operon.";
RL Mol. Gen. Genet. 259:105-114(1998).
DR EMBL: AF029224; AAC82542.1;
DR InterPro: IPR001457; Molybdopterin.
DR Pfam: PF001384; molybdopterin; 1.
DR DR PF01568; Molybdopterin; 1.
DR PROSITE: PS00551; MOLYBDOPTERIN_PROK_1; UNKNOWN.1.
SQ SEQUENCE 1224 AA; 138358 MW; 978D33279862C820 CRC64;

Query Match 50.0%; Score 7; DB 2; Length 1224;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VVDRDYT 9
DB 875 VVDRDYT 881

RESULT 5
O30293 ID O30293 PRELIMINARY; PRT; 107 AA.
AC O30293;1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF2377.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterdon S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;

```


RT "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*.";
RL Nature 390:364-370(1997).
DR EMBL: AE001112; AAB91292.1; -.
DR TIGR: AF2377; -.
DR InterPro: IPR003731; DUF153.
DR Pfam: PF02579; DUF153; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 107 AA; 11614 MW; F6CB538BDF6D9B4B CRC64;

Query Match 42.9%; Score 6; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGEVKV 14
| | | | |
Db 22 TGEVKV 27

RESULT 6
O42461 PRELIMINARY; PRT; 113 AA.
ID O42461
AC O42461;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PREPROGASTRIN.
OS Squalus acanthias (Spiny dogfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Squaloidei; Squalidae; Squalus.
OX NCBI_TaxID=7797;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=SPIRAL INTESTINE;
RX MEDLINE=97439845; PubMed=9294191;
RA Johnsen A.H., Joenson L., Rourke I.J., Rehfeld J.F.;
RT "Elasmobranchs express separate cholecystokinin and gastrin genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:10221-10226(1997).
DR EMBL: 297360; CAB10586.1; -.
DR InterPro: IPR001651; Gastrin.
DR Pfam: PF00918; Gastrin; 1.
DR PROSITE; PS00259; GASTRIN; 1.
DR SMART; SM00029; GASTRIN; 1.
FT CHAIN 53 102 GASTRIN-49.
FT CHAIN 85 102 GASTRIN-17.
FT CHAIN 94 102 GASTRIN-8.
SQ SEQUENCE 113 AA; 12582 MW; 3AECDA1ACC551565 CRC64;

Query Match 42.9%; Score 6; DB 13; Length 113;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DRDVTG 10
| | | | |
Db 92 DRDVTG 97

RESULT 7
Q9IBH9 PRELIMINARY; PRT; 113 AA.
ID Q9IBH9
AC Q9IBH9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GASTRIN.
OS Squalus acanthias (Spiny dogfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Squaloidei; Squalidae; Squalus.
OX NCBI_TaxID=7797;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=SPIRAL INTESTINE;
RX MEDLINE=97439845; PubMed=9294191;
RA Johnsen A.H., Joenson L., Rourke I.J., Rehfeld J.F.;
RT "Elasmobranchs express separate cholecystokinin and gastrin genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:10221-10226(1997).
DR EMBL: 297372; CAB96790.1; -.
DR EMBL: 297373; CAB96790.1; JOINED.
DR InterPro: IPR001651; Gastrin.
DR Pfam: PF00918; Gastrin; 1.
DR SMART; SM00029; GASTRIN; 1.
DR PROSITE; PS00259; GASTRIN; 1.
SQ SEQUENCE 113 AA; 12597 MW; 94E2241ACC551579 CRC64;

Query Match 42.9%; Score 6; DB 13; Length 113;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DRDVTG 10
| | | | |
Db 92 DRDVTG 97

RESULT 8
O70107 PRELIMINARY; PRT; 132 AA.
ID O70107
AC O70107;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE REGULATORY PROTEIN.
GN MERR.
OS Bacillus cereus.
OG Plasmid pKLH301.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN3-TYPE, TN1546-LIKE;
RA Minakhin L.S.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN3-TYPE, TN1546-LIKE;
RX MEDLINE=98195721; PubMed=9534232;
RA Bogdanova E.S., Bass I.A., Minakhin L.S., Petrova M.A., Mindlin S.Z.,
RA Volodin A.A., Kalyaeva E.S., Tiedge G.M., Hobman J.L., Brown N.L.,
RA Nikifirov V.G.;
RT "Horizontal spread of mer operons among gram-positive bacteria in natural environments.";
RL Microbiology 144:609-620(1998).
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL REGULATORS.

DR EMBL: Y09027; CAA70242.1; -.
DR EMBL: Y09024; CAA70225.1; -.
DR InterPro: IPR000551; HTH_MerR.
DR Pfam: PF00376; merR; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR; Transcription regulation.
KW DNA-binding; Plasmid; Transcription regulation.
SQ SEQUENCE 132 AA; 15988 MW; E6628416136E2A94 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GYVDRD 7
| | | | |
Db 70 GYVDRD 75

```

RESULT 9
O86214 PRELIMINARY; PRT; 132 AA.
AC O86214; 1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE REGULATORY PROTEIN.
GN MERR.
OS Exiguobacterium sp.
OC Plasmid pKHL3.
OC Bacillus/Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Exiguobacterium.
OX NCBI_TaxID=44751;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN-TC38-2B;
RC MEDLINE=98195721; PubMed=9534232;
RX Bogdanova E.S., Bass I.A., Minakhin L.S., Petrova M.A., Mindlin S.Z.,
RA Volodin A.A., Kalyaeva E.S., Tiedge G.M., Hobman J.L., Brown N.L.,
RA Nikiforov V.G.;
RT "Horizontal spread of mer operons among gram-positive bacteria in
natural environments.";
RL Microbiology 144:609-620(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN-TC38-2B;
RC STRAIN-TC38-2B;
RA Minakhin L.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN-TC38-2B;
RC STRAIN-TC38-2B;
RA Minakhin L.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
DR EMBL; X99457; CAA67818.1; -.
DR InterPro; IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTHMERR.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW DNA-binding; Plasmid; Transcription regulation.
SQ SEQUENCE 132 AA; 15960 MW; 6698FBF1FB95B635 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVVD RD 7
DB 70 GVVD RD 75

RESULT 10
O9F4C7 PRELIMINARY; PRT; 132 AA.
AC O9F4C7; 2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE REGULATORY PROTEIN.
GN MERR.
OS Bacillus licheniformis.
OC Bacillus/Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN-FA6-12;
RC "Bogdanova E.S., Minakhin L.S., Bass I.A., Hobman J.L., Brown N.L.,

```

```

RA Nikiforov V.G.;
RT "Mercury transposons in Gram-positive bacteria in natural
environments.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
DR EMBL; Y10855; CAA71810.1; -.
DR InterPro; IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 132 AA; 15929 MW; F7A88B1544DCAB28 CRC64;

```

```

Query Match 42.9%; Score 6; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVVD RD 7
DB 70 GVVD RD 75

```

```

RESULT 11
O9F4C0 PRELIMINARY; PRT; 132 AA.
AC O9F4C0;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE REGULATORY PROTEIN.
GN MERR.
OS Bacillus macroides.
OG Plasmid pKHL305.
OC Bacillus/Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=33935;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-TC47-5;
RA Bogdanova E.S., Minakhin L.S., Bass I.A., Hobman J.L., Brown N.L.,
RA Nikiforov V.G.;
RT "Mercury transposons in Gram-positive bacteria in natural
environments.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MERR FAMILY OF TRANSCRIPTIONAL
REGULATORS.
DR EMBL; Y09906; CAA71034.1; -.
DR InterPro; IPR000551; HTH_Merr.
DR Pfam; PF00376; merr; 1.
DR PRINTS; PR00040; HTHMERR.
DR SMART; SM00422; HTH_MERR; 1.
DR PROSITE; PS00552; HTH_MERR_FAMILY; 1.
KW DNA-binding; Plasmid; Transcription regulation.
SQ SEQUENCE 132 AA; 15919 MW; 9B7A1C32B558EB7D CRC64;

```

```

Query Match 42.9%; Score 6; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVVD RD 7
DB 70 GVVD RD 75

```

```

RESULT 12
O54602 PRELIMINARY; PRT; 206 AA.
ID O54602
AC O54602;
DT 01-JUN-1998 (TREMblrel. 06, Created)

```

DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE ORF H1613.
 OS Halobacterium sp. (strain NRC-1).
 OG Plasmid pNRC100.
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=88201675; PubMed=3448465;
 RA Dassarma S., Damerval T., Jones J.G., Tandeau de Marsac N.;
 RT "A plasmid-encoded gas vesicle protein gene in a halophilic
 archaeobacterium.";
 RL Mol. Microbiol. 1:365-370(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=90016863; PubMed=2552415;
 RA Jones J.G., Hackett N.R., Halladay J.T., Scothorn D.J., Yang C.F.,
 NG W.L., Dassarma S.;
 RT "Analysis of insertion mutants reveals two new genes in the pNRC100
 gas vesicle gene cluster of Halobacterium halobium.";
 RL Nucleic Acids Res. 17:7785-7793(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=91267967; PubMed=2050644;
 RA Ng W.L., Kothakota S., Dassarma S.;
 RT "Structure of the gas vesicle plasmid in Halobacterium halobium
 inversion isomers, inverted repeats, and insertion sequences.";
 RL J. Bacteriol. 173:3933-3933(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=91323716; PubMed=1864501;
 RA Jones J.G., Young D.C., Dassarma S.;
 RT "Structure and organization of the gas vesicle gene cluster on the
 Halobacterium halobium plasmid pNRC100.";
 RL Gene 102:117-122(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=93012964; PubMed=1398080;
 RA Halladay J.T., Ng W.L., Dassarma S.;
 RT "Genetic transformation of a halophilic archaeobacterium with a gas
 vesicle gene cluster restores its ability to float.";
 RL Gene 119:131-136(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=93139036; PubMed=8423144;
 RA Halladay J.T., Jones J.G., Lin F., MacDonald A.B., Dassarma S.;
 RT "The rightward gas vesicle operon in Halobacterium plasmid pNRC100:
 identification of the gvpA and gvpC gene products by use of antibody
 probes and genetic analysis of the region downstream of gvpC.";
 RL J. Bacteriol. 175:684-692(1993).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=93327890; PubMed=8335077;
 RA Dassarma S.;
 RT "Identification and analysis of the gas vesicle gene cluster on an
 unstable plasmid of Halobacterium halobium.";
 RL Experientia 49:482-486(1993).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=93328662; PubMed=8335618;
 RA Ng W.L., Dassarma S.;
 RT "Minimal replication origin of the 200-kilobase Halobacterium plasmid
 pNRC100."

RL J. Bacteriol. 175:4584-4596(1993).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=9505934; PubMed=8002589;
 RA Dassarma S., Arora P., Lin F., Molinari E., Yin L.R.;
 RT "Wild-type gas vesicle formation requires at least ten genes in the
 gvp gene cluster of Halobacterium halobium plasmid pNRC100.";
 RL J. Bacteriol. 176:7646-7652(1994).
 RN [10]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX NG W.L., Arora P., Dassarma S.;
 RA Syst. Appl. Microbiol. 16:560-568(1994).
 RN [11]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX Dassarma S.;
 RA (in) Robb F.T., Place A.R., Sowers K.R., Schreier H.J., Dassarma S.,
 RL Fleischmann E.M. (eds.);
 RL Archaea:
 RL A laboratory manual - halophiles, pp.253-255,
 RL Cold Spring Harbor Laboratory Press, New York (1995).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX Dassarma S., Arora P.;
 RL FEMS Microbiol. Lett. 153:1-10(1997).
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX NG W.L., Ciuffo S.A., Smith T.M., Bumgarner R.E., Loretz C., Baskin D.,
 RA Faust J., Seto J., Slagel J., Hood L., Dassarma S.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF016485; AAC82925.1; -;
 DR EMBL; AF016485; AAC82852.1; -;
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 206 AA; 22044 MW; 3887A5929EA5992B CRC64;
 QY 3 VVDRDY 8
 Db 139 VVDRDY 144
 PRT; 279 AA.
 RESULT 13
 Q9EVM9 PRELIMINARY;
 ID Q9EVM9 PRT; 279 AA.
 AC Q9EVM9;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 31.3 KDA PROTEIN.
 OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=316;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A15;
 RA Desnoues N., Lin M., Elmerich C.;
 RT "Organisation of nif genes in Pseudomonas stutzeri A15, a rice
 endophyte."

Query Match 42.9%; Score 6; DB 1; Length 206;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RL Submitted (JUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ297529; CAC03733.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 279 AA; 31298 MW; 91473D0FE3B31491 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 279;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVVD RD 7
 Db 24 GVVD RD 29
 |||||

RESULT 14

O30655 PRELIMINARY; PRT; 287 AA.
 ID O30655;
 AC O30655;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last annotation update)
 DE HYPOTHETICAL 32.2 KDa PROTEIN.
 DE
 OS Azotobacter vinelandii;
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Azotobacter;
 OC NCBI_TaxID=354;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=TRANS;
 RA Dean D.R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF014048; AAB6328.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 287 AA; 32170 MW; 6541031DD01B7D3D CRC64;

Query Match 42.9%; Score 6; DB 2; Length 287;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVVD RD 7
 Db 29 GVVD RD 34
 |||||

RESULT 15

O9CKLO PRELIMINARY; PRT; 317 AA.
 ID O9CKLO;
 AC O9CKLO;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE TAL_1 OR PM1602.
 GN TAL_1 OR PM1602.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella;
 OC NCBI_TaxID=747;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=PM70;
 RC MEDLINE=21145866; PubMed=11248100;
 RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RA "Complete genomic sequence of Pasteurella multocida pm70.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 RL EMBL; AE006198; AAK03686.1; -
 DR InterPro; IPR001585; Transaldolase.
 DR Pfam; PF00923; Transaldolase; 1.
 DR PROSITE; PS00958; TRANSALDOLASE_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 317 AA; 35007 MW; 4DCA0CFED73458DA CRC64;

Query Match 42.9%; Score 6; DB 2; Length 317;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YTG EVK 13
 Db 265 YTG EVK 270
 |||||

Search completed: January 31, 2002, 13:37:50
 Job time: 175 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:17 ; Search time 78.64 seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-79
Perfect score: 14
Sequence: 1 AGVDRDVTGEVKV 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	147	S38189	dUTP pyrophosphatase
2	7	50.0	361	T21136	hypothetical prote
3	7	50.0	749	T08101	outer dynein arm d
4	6	42.9	107	A69547	conserved hypothet
5	6	42.9	123	C71855	ATP synthase F1, c
6	6	42.9	124	C64661	ATP synthase F1, s
7	6	42.9	132	A32239	mercuric resistanc
8	6	42.9	132	T45509	regulatory protein
9	6	42.9	132	T44501	merri protein [imp
10	6	42.9	185	E64126	hypothetical prote
11	6	42.9	206	T08285	hypothetical prote
12	6	42.9	317	S40535	transaldolase (EC
13	6	42.9	317	H85480	transaldolase B [i
14	6	42.9	355	T40385	hypothetical prote
15	6	42.9	376	H82308	oxaloacetate decar
16	6	42.9	379	F83496	probable acyl-CoA
17	6	42.9	387	T51225	related to INORGAN
18	6	42.9	394	E83213	probable acyl-CoA
19	6	42.9	406	T45863	hypothetical prote
20	6	42.9	438	B83295	hypothetical prote
21	6	42.9	491	G81892	anthranilate synth
22	6	42.9	491	E81132	anthranilate synth
23	6	42.9	621	A75101	aldehyde--ferredox
24	6	42.9	625	G71072	aldehyde--ferredox
25	6	42.9	802	C83588	probable hydroxama
26	6	42.9	811	PN0589	connectin 1 - chic
27	6	42.9	949	D82293	isoleucyl-tRNA syn
28	6	42.9	1434	C82923	DNA-directed RNA p
29	6	42.9	1983	G86643	hypothetical prote

30	6	42.9	3071	2	T45584	hypothetical prote
31	6	42.9	4162	2	T42633	connectin/titin -
32	6	42.9	26926	1	I38344	titin, cardiac mus
33	5	35.7	60	2	F69270	ferredoxin (fdx-2)
34	5	35.7	61	2	E64575	hypothetical prote
35	5	35.7	67	2	A69342	hypothetical prote
36	5	35.7	77	2	T07551	hypothetical prote
37	5	35.7	98	2	D30338	exogenous DNA-bind
38	5	35.7	100	2	T48765	hypothetical prote
39	5	35.7	103	2	F75334	probable dioxygena
40	5	35.7	107	2	T12133	leghemoglobin - fa
41	5	35.7	110	2	A37370	lamellar cuticular
42	5	35.7	110	2	B37370	lamellar cuticular
43	5	35.7	111	1	KVMSCL	Ig kappa chain V r
44	5	35.7	114	1	DCL8HB	histidine decarbox
45	5	35.7	116	2	C75451	hypothetical prote

ALIGNMENTS

RESULT 1
S38189
dUTP pyrophosphatase (EC 3.6.1.23) precursor, mitochondrial - yeast (Saccharomyces ce
N;Alternate names: protein YBR1705; protein YBR252W
C;Species: Saccharomyces cerevisiae
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C;Accession: S38189; S46133; S39578; S36425
R;Doignon, F.; Biteau, N.; Aigle, M.; Crouzet, M.
Yeast 9, 1131-1137, 1993
A;Title: The complete sequence of a 6794 bp segment located on the right arm of chrom
A;Reference number: S38185; MUID:94078675
A;Accession: S38189
A;Molecule type: DNA
A;Residues: 1-147 <DOI>
A;Cross-references: GB:L20296; NID:g311101; PIDN:AAA65611.1; PID:g311106
R;Aigle, M.; Baclet, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Doignon, F.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45940
A;Accession: S46133
A;Molecule type: DNA
A;Residues: 1-147 <AIG>
A;Cross-references: EMBL:Z36121; NID:g536670; PIDN:CAA85215.1; PID:g536671; MIPS:YBR2
R;Gadsden, M.H.; McIntosh, E.M.; Game, J.C.; Willson, P.J.; Haynes, R.H.
EMBO J. 12, 4425-4431, 1993
A;Title: dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae.
A;Reference number: S39578; MUID:94038925
A;Accession: S39578
A;Molecule type: DNA
A;Residues: 1-9, 'N', 11-147 <GAD>
A;Cross-references: EMBL:X74263; NID:g398183; PIDN:CAA52322.1; PID:g398184
R;Gadsden, M.; McIntosh, E.; Game, J.; Willson, P.; Haynes, R.
submitted to the EMBL Data Library, July 1993
A;Description: dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisia
A;Reference number: S36425
A;Accession: S36425
A;Molecule type: DNA
A;Residues: 1-9, 'N', 11-95, 'V', 97-138, 'R', 140-146, 'K' <GA2>
A;Cross-references: EMBL:X74263
C;Genetics:
A;Gene: SGD:DUT1
A;Cross-references: SGD:S0000456; MIPS:YBR252W
A;Map position: 2R
A;Genome: nuclear
C;Superfamily: retroviral proteinase
C;Keywords: hydrolase; mitochondrion; nucleotide metabolism; proteinase
F:1-15/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:16-147/Product: dUTP pyrophosphatase #status predicted <MAT>

Query Match 100.0%; Score 14; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 6e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVVDRTDYTGKVK 14
| | | | | | | | | | | | | | | | | |
Db 81 AGVVDRTDYTGKVK 94

RESULT 2

T21136
hypothetical protein F20C5.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T21136
R:Matthews, P.
Submitted to the EMBL Data Library, December 1995
A:Reference number: Z19381
A:Accession: T21136
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-361 <WTL>
A:Cross-references: EMBL:268161; PIDN:CAA92297.1; GSPDB:GN00022; CESP:F20C5.4
A:Experimental source: clone F20C5
C:Genetics:
A:Gene: CESP:F20C5.4
A:Map position: 4
A:Introns: 20/1; 65/1; 109/3; 175/1; 226/3; 301/1
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YNR048w

Query Match 50.0%; Score 7; DB 2; Length 361;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DYTGEVK 13

Db 107 DYTGEVK 113
| | | | | | | | | | | | | | | | | |

RESULT 3

T08101
outer dynein arm docking complex protein ODA3 - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999
C:Accession: T08101
R:Koutoulis, A.; Pazour, G.J.; Wilkerson, C.G.; Inaba, K.; Sheng, H.; Takada, S.; Witman
J. Cell Biol. 137, 1069-1080, 1997
A:Title: The Chlamydomonas reinhardtii ODA3 gene encodes a protein of the outer dynein arm
A:Reference number: Z16354; MUID:97311076
A:Accession: T08101
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-749 <KOU>
A:Cross-references: EMBL:AF001309; NID:g2150040; PIDN:AAC49732.1; PID:g2150041
C:Genetics:
A:Gene: ODA3
C:Function:
A:Description: may play an important role in the precise positioning of the outer dynein

Query Match 50.0%; Score 7; DB 2; Length 749;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVVDRDY 8

Db 722 GVVDRDY 728
| | | | | | | | | | | | | | | | | |

RESULT 4

A69347
conserved hypothetical protein AF2377 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: A69347

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69347; MUID:98049343
A:Accession: A69347
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-107 <KLE>
A:Cross-references: GB:AE001112; GB:AE000782; NID:g2689435; PIDN:AAB91292.1; PID:g265
C:Superfamily: conserved hypothetical protein M70580

Query Match 42.9%; Score 6; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGEVKV 14

Db 22 TGEVKV 27
| | | | | | | | | | | | | | | | | |

RESULT 5

C71855
ATP synthase F1, chain epsilon - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Jun-1999
C:Accession: C71855
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
Ives, G.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: C71855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-123 <ARN>
A:Cross-references: GB:AE001533; GB:AE001439; NID:g4155636; PIDN:AAD06638.1; PID:g415
A:Experimental source: strain J99
C:Genetics:
A:Gene: atpC
C:Superfamily: H⁺-transporting ATP synthase epsilon chain

Query Match 42.9%; Score 6; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YTGEVK 13

Db 16 YTGEVK 21
| | | | | | | | | | | | | | | | | |

RESULT 6

C64661
ATP synthase F1, subunit epsilon - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 22-Jun-1999
C:Accession: C64661
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McK
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: C64661
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A;Residues: 1-124 <TOM>
A;Cross-references: GB:AE000619; GB:AE000511; NID:g2314276; PIDN:AAD08173.1; PID:g231428
C;Genetics:
A;Start codon: GNG
C;Superfamily: H⁺-transporting ATP synthase epsilon chain

Query Match 42.9%; Score 6; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YTEVK 13
Db 17 YTEVK 22

RESULT 7
A32239
mercuric resistance operon regulatory protein - Bacillus sp.
C;Species: Bacillus sp.
C;Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 15-Oct-1999
C;Accession: A32239; A32227
R;Helmann, J.D.; Wang, Y.; Mahler, I.; Walsh, C.T.
J. Bacteriol. 171, 222-229, 1989
A;Title: Homologous metalloregulatory proteins from both gram-positive and gram-negative
A;Reference number: A32239; MUID:89123021
A;Accession: A32239
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-132 <HEL>
R;Wang, Y.; Moore, M.; Levinson, H.S.; Silver, S.; Walsh, C.; Mahler, I.
J. Bacteriol. 171, 83-92, 1989
A;Title: Nucleotide sequence of a chromosomal mercury resistance determinant from a Bacillus
A;Reference number: A32227; MUID:89123092
A;Accession: A32227
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-44, 'T', 46-132 <WAN>
A;Cross-references: GB:M22708; NID:g143188; PIDN:AAA83973.1; PID:g1129093
C;Superfamily: transcription repressor glrR
C;Keywords: DNA binding; transcription regulation

Query Match 42.9%; Score 6; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVDRD 7
Db 70 GVDRD 75

RESULT 8
T45509
regulatory protein [imported] - Exiguobacterium sp. plasmid pKLH3
C;Species: Exiguobacterium sp.
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45509
R;Bogdanova, E.S.; Bass, I.A.; Minhakhin, L.S.; Petrova, M.A.; Mindlin, S.Z.; Volodin, A.
Microbiology 144, 609-620, 1998
A;Title: Horizontal spread of mer operons among Gram-positive bacteria in natural environment
A;Reference number: 222993; MUID:98195721
A;Accession: T45509
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-132 <BO>
A;Cross-references: EMBL:X99457; NID:g3413183; PIDN:CAA67818.1; PID:g3413184
A;Experimental source: strain TC38-2b
C;Genetics:
A;Gene: merr
A;Genome: plasmid pKLH3
C;Function:
A;Description: involved in mercury resistance

C;Superfamily: transcription repressor glrR

Query Match 42.9%; Score 6; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVDRD 7
Db 70 GVDRD 75

RESULT 9
T44501
merR protein [imported] - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C;Accession: T44501
R;Narita, M.; Koizumi, T.; Huang, C.; Endo, G.
submitted to the EMBL Data Library, March 1999
A;Description: Broad-spectrum mercury resistance and its genetic characterization of
n.
A;Reference number: Z22785
A;Accession: T44501
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-132 <NAR>
A;Cross-references: EMBL:AB024961; PIDN:BA086113.1
A;Experimental source: isolate Mersaru
C;Genetics:
A;Gene: merR
C;Superfamily: transcription repressor glrR

Query Match 42.9%; Score 6; DB 2; Length 132;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GVDRD 7
Db 70 GVDRD 75

RESULT 10
E64126
hypothetical protein H11488 - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C;Accession: E64126
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: E64126
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-185 <TIGR>
A;Cross-references: GB:U32826; GB:L42023; NID:g1574322; PIDN:AAC23133.1; PID:g1574324
C;Superfamily: Haemophilus influenzae hypothetical protein H11488

Query Match 42.9%; Score 6; DB 2; Length 185;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGEVK 14
Db 77 TGEVK 82

```
RESULT 11
T08285
hypothetical protein H0748 [imported] - Halobacterium sp. (strain NRC-1) plasmid pNRC100
N:Alternate names: hypothetical protein H1613
C:Species: Halobacterium sp.
A:Variety: strain NRC-1
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 03-Nov-2000
C:Accession: T08285; T08358
R:Ng, W.V.; Cluifo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baskin, D.; Faust, J.; Hall, B.;
Genome Res. 8, 1131-1141, 1998
A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid or
A:Reference number: Z16408; MUID:99063795
A:Accession: T08285
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-206 <NGW>
A:CROSS-references: EMBL:AF016485; NID:g2822278; PID:g2822346; HALOSP:H0748
A:Experimental source: strain NRC-1
A:Genetics: COP1
A:Accession: T08358
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-206 <DAS>
A:CROSS-references: EMBL:AF016485; NID:g2822278; PID:g2822419; HALOSP:H1613
A:Experimental source: strain NRC-1
A:Genetics: COP2
A:Genetics: <COP1>
A:Gene: HALOSP:H0748
A:Genome: plasmid pNRC100
C:Genetics: <COP2>
A:Gene: HALOSP:H1613
A:Genome: plasmid pNRC100
C:Superfamily: Halobacterium plasmid pNRC100 hypothetical protein H0748

Query Match 42.9%; Score 6; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 22; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 3 VVDRDY 8
Db 139 VVDRDY 144

RESULT 12
S40535
transaldolase (EC 2.2.1.2) B - Escherichia coli
C:Species: Escherichia coli
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Jun-2000
C:Accession: S40535; H64720
R:Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu
submitted to the EMBL Data Library, December 1992
A:Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A:Reference number: S40531
A:Accession: S40535
A:Molecule type: DNA
A:Residues: 1-317 <YUR>
A:CROSS-references: EMBL:D10483; NID:g216434; PIDN:BAA01290.1; PID:g216439
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H64720
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-317 <BLAT>
A:CROSS-references: GB:AE000111; GB:U00096; NID:g1786181; PIDN:AAC73119.1; PID:g1786189;
A:Experimental source: strain K-12, substrain MGL655
A:Genetics:
A:Gene: taib
A:Function:
A:Description: catalyzes the reversible transfer of a three-carbon ketol unit from sedoh
```

```
C:Superfamily: human transaldolase
C:Keywords: transferase
F:132/Active site: Lys #status predicted

Query Match 42.9%; Score 6; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 33; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 8 YTGVEK 13
Db 265 YTGVEK 270

RESULT 13
H85480
transaldolase B [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: H85480
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85480
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-317 <STO>
A:CROSS-references: GB:AE005174; NID:g12512683; PIDN:AAG54308.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: taib
C:Superfamily: human transaldolase

Query Match 42.9%; Score 6; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 33; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 8 YTGVEK 13
Db 265 YTGVEK 270

RESULT 14
T40385
hypothetical protein SPBC3E7.11c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: T40385
R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Brown, D.; Churcher, C.M.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21924
A:Accession: T40385
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-355 <LYN>
A:CROSS-references: EMBL:AL023534; PIDN:CAA19014.1; GSPDB:GN000067; SPDB:SPBC3E7.11c
A:Experimental source: strain 972h-; cosmid c3E7
C:Genetics:
A:Gene: SPDB:SPBC3E7.11c
A:Map position: 2
A:Introns: 30/2; 54/3
A:Superfamily: dnaJ amino-terminal homology
F:9-75/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 42.9%; Score 6; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 36; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

Qy 3 VVDRDY 8
```



```

Db      5 VVDRDY 10
|||||
RESULT 15
H82308
oxaloacetate decarboxylase, beta chain VC0551 [imported] - Vibrio cholerae (strain N1696)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: H82308
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: H82308
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <HEI>
A:Cross-references: GB:AF004141; GB:AF003852; NID:g9654976; PIDN:AAF93719.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0551
A:Map position: 1
C:Superfamily: Propionigenium modestum methylmalonyl-CoA decarboxylase beta chain

```

```

Query Match      42.9%; Score 6; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGVVDR 6
      |||||
Db      244 AGVVDR 249

```

Search completed: January 31, 2002, 13:20:19
Job time: 120 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:13 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-79

Perfect score: 14

Sequence: 1 AGVWDRDYTGKVKV 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	19	AAW72855 yeast dUTPase urid
2	7	50.0	102	21	AA840346 Human OREF ORF110
3	7	50.0	1224	18	AAW26735 Staphylococcus car
4	6	42.9	44	18	AAW27783 UDP-N-acetylglucos
5	6	42.9	124	18	AAW20256 H. pylori outer me
6	6	42.9	252	20	AAW96156 Staphylococcus aureu
7	6	42.9	421	20	AAW96155 Staphylococcus aureu
8	6	42.9	421	21	AAW70130 Staphylococcus aur
9	6	42.9	421	21	AAW70137 Staphylococcus aur
10	6	42.9	867	20	AAW34508 Porphyromonas ging
11	6	42.9	875	20	AAW34381 Porphyromonas ging

12	6	42.9	1462	20	AAW01519 A carcinogenesis-i
13	6	42.9	26926	22	AAU05396 Human titin (conne
14	5	35.7	9	18	AAW25728 Labelled cholecyst
15	5	35.7	9	22	AAW37654 Cholecystokinin pe
16	5	35.7	9	22	AAW37655 Cholecystokinin pe
17	5	35.7	32	20	AAW82454 Onchocerca sp Cu/2
18	5	35.7	32	22	AAW19992 Peptide #6426 enco
19	5	35.7	32	22	AAW33697 Peptide #7734 enco
20	5	35.7	47	15	AAW56632 Tyrosine activatio
21	5	35.7	48	15	AAW56631 Tyrosine activatio
22	5	35.7	63	21	AAW18871 zea mays protein f
23	5	35.7	84	19	AAW71635 Omega-cyclohexane
24	5	35.7	89	21	AAW25845 Arabidopsis thalia
25	5	35.7	101	21	AAW25844 Arabidopsis thalia
26	5	35.7	102	22	AAW21184 Peptide #7618 enco
27	5	35.7	102	22	AAW37399 Peptide #11436 enc
28	5	35.7	104	22	AAW92501 Nicotiana tabacum
29	5	35.7	115	21	AAW61462 Arabidopsis thalia
30	5	35.7	120	21	AAW04411 Arabidopsis thalia
31	5	35.7	132	21	AAW16405 pinus radiata PAL
32	5	35.7	133	19	AAW77684 4-Methyl-5(beta-hy
33	5	35.7	143	18	AAW09017 immunogenic type F
34	5	35.7	146	21	AAW61461 Arabidopsis thalia
35	5	35.7	149	18	AAW28215 Amino acid sequenc
36	5	35.7	149	22	AAW17893 Peptide #4327 enco
37	5	35.7	149	22	AAW30402 Peptide #4439 enco
38	5	35.7	149	22	AAW05546 Peptide #4228 enco
39	5	35.7	152	21	AAW52547 Murine Ig-beta/mul
40	5	35.7	155	21	AAW41380 Arabidopsis thalia
41	5	35.7	162	22	AAW92518 C glutamicum prote
42	5	35.7	162	22	AAW79203 Corynebacterium gl
43	5	35.7	163	21	AAW33164 zea mays protein f
44	5	35.7	165	21	AAW26360 Arabidopsis thalia
45	5	35.7	165	21	AAW32119 Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW72855
ID AAW72855 standard; Peptide; 14 AA.
XX
AC AAW72855;
XX
DT 01-MAR-1999 (first entry)
XX
DE Yeast dUTPase uridine-binding motif.
XX
KW Polymerase enhancing factor; PEF; dUTPase; PCR; amplification;
KW sequencing; replication.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9842860-A1.
XX
PD 01-OCT-1998.
XX
PF 20-MAR-1998; 98WO-US05497.
XX
PR 24-OCT-1997; 97US-0957709.
PR 21-MAR-1997; 97US-0822774.
XX
XX (STRA-) STRATAGENE.
XX
XX Hansen CJ, Hogrefe H;
XX WPI; 1998-542284/46.
XX
XX Polymerase enhancing factor proteins, extracts and complexes -
XX improve the polymerisation activity of nucleic acid polymerases, for
XX use in amplification, sequencing and replication

PS Claim 71; Page 47; 161pp; English.

XX This is the uridine-binding motif of yeast dUTPase. Sequences
 CC are provided (see AAW72849-57) of the uridine-binding motifs of the
 CC dUTPases and dCTP deaminases of *Pyrococcus furiosus* (see also
 CC AAW72847), *Methanococcus jannaschii*, *Desulfurolobus ambivalens*,
 CC *Escherichia coli*, yeast, human and herpesvirus; a consensus (see
 CC AAW72848) is also provided. A claimed method of enhancing a nucleic
 CC acid polymerase reaction comprises performing the reaction in the
 CC presence of one or more of the following: a polymerase enhancing
 CC factor (PEF), a dUTPase, a protein that turns-over dUTP and a
 CC protein having one or more of the sequences provided in AAW72848-57.
 CC A claimed protein having PEF activity comprises one or more of
 CC sequences given in AAW72848-57. Kits are provided for replicating
 CC nucleic acids, for site-directed mutagenesis, for nucleic acid
 CC sequencing or for amplification (preferably PCR or RT-PCR).

XX Sequence 14 AA;

SO Query Match 100.0%; Score 14; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.5e-08;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVVDRDYTGCVK 14
 Db 1 agvdrdytgevk 14

RESULT 2
 AAB40346
 ID AAB40346 standard; Protein: 102 AA.
 AC AAB40346;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF110 polypeptide sequence SEQ ID NO:220.
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

XX Homo sapiens.
 XX
 XX W0200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Leach M;
 XX WPI: 2000-602362/57.
 XX DR N-PSDB; AAC74555.

XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 547-548; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nootropic; immunoprotective;
 CC osteopathic; anticonvulsant; antiarthritic; neurosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy.
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 102 AA;

SO Query Match 50.0%; Score 7; DB 21; Length 102;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGVVDRD 7
 Db 45 agvvdrd 51

RESULT 3
 AAW26735
 ID AAW26735 standard; Protein: 1224 AA.
 XX AAW26735;
 AC AAW26735;
 DT 11-MAY-1998 (first entry)
 DE Staphylococcus carnosus nitrate reductase NarG subunit.
 KW Nitrate reductase; NarG; nitrite reductase; nitrate reduction;
 KW nitrite reduction; pollutant; biotransformation; water treatment;
 KW vegetable.
 XX
 XX Staphylococcus carnosus DSM 10563.
 OS
 XX EP805205-A1.
 PN
 XX
 PD 05-NOV-1997.
 XX
 PF 02-MAY-1996; 96EP-0201224.
 XX
 PR 02-MAY-1996; 96EP-0201224.
 XX
 XX (NEST) SOC PROD NESTLE SA.
 PA
 XX Fast B, Gaier W, Goetz F, Lindgren P, Neubauer H;
 PI Pantel I;
 PI WPI: 1997-529059/49.
 DR N-PSDB; AAT99449.
 XX

PT Recombinant *S. carnosus* nitrate and nitrite reductase proteins -
 PT useful for reducing nitrate and nitrite in nitrate-polluted material
 XX
 PS Claim 1; Page 29-32; 66pp; English.

XX This protein comprises the G subunit of nitrate reductase NARGHJ
 CC encoded by nucleotides 4140-7814 of an isolated nitrite reductase
 CC and nitrate reductase gene region (see AAT99449) of *Staphylococcus*
 CC *carnosus*. The invention relates to recombinant proteins (see
 CC AAW26733-49) involved in the reduction of nitrate and nitrite in *S.*
 CC *carnosus*, and the use of these proteins, or cells encoding them,
 CC in free or immobilised form to reduce nitrate and nitrite in
 CC water, vegetables and other nitrate and/or nitrite polluted
 CC materials.

XX Sequence 1224 AA;

Query Match 50.0%; Score 7; DB 18; Length 1224;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VVDRDYT 9
 Db 875 vvdrdyt 881
 |||||

RESULT 4
 AAW27783
 ID AAW27783 standard; Protein; 44 AA.

XX AAW27783;
 XX 21-JUL-1998 (first entry)

DE UDP-N-acetylglucosamine 1-carboxyvinyltransferase.

XX *Staphylococcus aureus* protein; ribozyme; antisense sequence; control;
 KW *Staphylococcus aureus* gene; regulatory element; bacterial gene expression;
 KW vaccine; *Staphylococcus aureus* infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome.

XX *Staphylococcus aureus*.

XX WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI; 1997-424969/39.

DR N-PSDB; AAT83752.

XX Novel polypeptide(s) from *Staphylococcus aureus* strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against *S.*
 PT *aureus* infection

PS Claim 6; Page 272; 989pp; English.

XX The present sequence represents a *Staphylococcus aureus* protein, that,
 CC based on homology with an *Acinetobacter calcoaceticus* protein, is
 CC believed to be a UDP-N-acetylglucosamine 1-carboxyvinyltransferase
 CC (enolpyruvate transferase, UDP-N-acetylglucosamine enolpyruvyl
 CC transferase). The DNA sequence was isolated from a library of clones of
 CC *S. aureus* WCUH 29 in *Escherichia coli*. The DNA sequence can be used in
 CC the construction of ribozymes and antisense sequences to control the

CC expression of *Staphylococcus* genes. The DNA sequence is also useful as
 CC a source of regulatory elements for the control of bacterial gene
 CC expression. The present protein may be used to produce vaccines to enable
 CC a host to produce specific antibodies with antibacterial action. These
 CC vaccines and antibodies would protect a host against infection by
 CC *S. aureus*, and conditions relating to *Staphylococcus aureus* infection,
 CC e.g. *Staphylococcus aureus* food poisoning, scaled skin syndrome, and toxic shock
 CC syndrome.

XX Sequence 44 AA;

Query Match 42.9%; Score 6; DB 18; Length 44;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGEVKV 14
 Db 13 tgekv 18
 |||||

RESULT 5
 AAW20256
 ID AAW20256 standard; Protein; 124 AA.

XX AAW20256;

XX 09-JUL-1997 (first entry)

XX H. pylori outer membrane protein 23531562.aa.

XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

XX *Helicobacter pylori*.

XX WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US09122.

XX 01-APR-1996; 96US-0630405.

XX 07-JUN-1995; 95US-0487032.

XX (ASTR) ASTRA AB.

XX Berglindh OT, Smith D, Mellgaard BL;

XX WPI; 1997-052306/05.

DR N-PSDB; AAT67471.

XX *Helicobacter pylori* nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent *H. pylori*
 PT infection, and to detect *Helicobacter*

PS Claim 56; Pages 461; 1481pp; English.

XX This sequence is a *H. pylori* outer membrane protein.
 CC The protein may be used in a vaccine to prevent or treat *H. pylori*
 CC infection or to identify *H. pylori* polypeptide binding compounds,
 CC useful as potential *H. pylori* life cycle activators or inhibitors.
 CC The genomic sequence of *H. pylori* (ATCC 53679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely *H. pylori* antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from *H. pylori* by PCR amplification for recombinant polypeptide
 CC production, e.g. in *E. coli* hosts.

XX SQ Sequence 124 AA;

Query Match 42.9%; Score 6; DB 18; Length 124;
Best Local Similarity 100.0%; Pred No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 YTGVEVK 13
Db 17 YTGVEK 22

RESULT 6
AAW96156
ID AAW96156 standard; Protein; 252 AA.

XX AC AAW96156;

XX DT 27-APR-1999 (first entry)

XX DE Staphylococcus aureus MurA protein fragment from open reading frame.

XX KW Staphylococcus aureus; murA; antibodies; antibiotics; therapy;

XX KW infection; UDP-N-acetylglucosamine enolpyruvyltransferase;

XX KW diagnosis; bactericidal; bacteriostatic; osteomyelitis;

XX KW septic arthritis; septic thrombophlebitis;

XX KW acute bacterial endocarditis; toxic shock syndrome;

XX KW scalded skin syndrome; food poisoning; treatment;

XX KW immunogen.

XX OS Staphylococcus aureus.

XX PN EP890644-A2.

XX PD 13-JAN-1999.

XX PF 01-JUL-1998; 98EP-0305253.

XX PR 10-JUL-1997; 97US-0052214.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PA (SMIK) SMITHKLINE BEECHAM PLC.

XX PI Wallis NG;

XX DR WPI; 1999-072880/07.

XX DR N-PSDB; AAX08914.

XX PT New Streptococcus aureus UDP-N-acetylglucosamine
enolpyruvyltransferase (murA) polypeptides and polynucleotides -
PT useful as diagnostic reagents and for prevention and treatment of
PT Streptococcus aureus infections

XX PS Claim 1; Page 34-35; 40pp; English.

XX CC MurA polypeptides and polynucleotides are useful for diagnosing
disease and susceptibility to disease by detecting mutations or
polymorphisms in the MurA gene or analysing for the presence or
amount of MurA polypeptide expressed in a patient. MurA
polypeptides and polynucleotides are useful for screening for
antagonists, agonists and drugs against infection from
micro-organisms. MurA agonists and antagonists are bacteriostatic and
bacteriocidal compounds which can be used in treatment to enhance
or block MurA activity, therefore treating bacterial infections,
especially by Staphylococcus aureus which causes bacteraemia in cancer
patients, osteomyelitis, septic arthritis, septic thrombophlebitis,
acute bacterial endocarditis, toxic shock syndrome, scalded skin
syndrome and Staphylococcal food poisoning. Epitopes of MurA
polypeptides and polynucleotides are useful immunogens for producing
anti-MurA antibodies for prevention of bacterial infections.

XX SQ Sequence 252 AA;

Query Match 42.9%; Score 6; DB 20; Length 252;
Best Local Similarity 100.0%; Pred No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 TGEVKV 14
Db 13 tgevkv 18

RESULT 7
AAW96155
ID AAW96155 standard; Protein; 421 AA.

XX AC AAW96155;

XX DT 27-APR-1999 (first entry)

XX DE Staphylococcus aureus MurA protein.

XX KW Staphylococcus aureus; murA; antibody; antibiotic; therapy;

XX KW infection; UDP-N-acetylglucosamine enolpyruvyltransferase;

XX KW diagnosis; bactericidal; bacteriostatic; osteomyelitis;

XX KW septic arthritis; septic thrombophlebitis; food poisoning;

XX KW acute bacterial endocarditis; toxic shock syndrome;

XX KW scalded skin syndrome; treatment; immunogen.

XX OS Staphylococcus aureus.

XX PN EP890644-A2.

XX PD 13-JAN-1999.

XX PF 01-JUL-1998; 98EP-0305253.

XX PR 10-JUL-1997; 97US-0052214.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PA (SMIK) SMITHKLINE BEECHAM PLC.

XX PI Wallis NG;

XX DR WPI; 1999-072880/07.

XX DR N-PSDB; AAX08913.

XX PT New Streptococcus aureus UDP-N-acetylglucosamine
enolpyruvyltransferase (murA) polypeptides and polynucleotides -
PT useful as diagnostic reagents and for prevention and treatment of
PT Streptococcus aureus infections

XX PS Claim 1; Page 32-33; 40pp; English.

XX CC MurA polypeptides and polynucleotides are useful for diagnosing
disease and susceptibility to disease by detecting mutations or
polymorphisms in the MurA gene or analysing for the presence or
amount of MurA polypeptide expressed in a patient. MurA
polypeptides and polynucleotides are useful for screening for
antagonists, agonists and drugs against infection from
micro-organisms. MurA agonists and antagonists are bacteriostatic and
bacteriocidal compounds which can be used in treatment to enhance
or block MurA activity, therefore treating bacterial infections,
especially by Staphylococcus aureus which causes bacteraemia in cancer
patients, osteomyelitis, septic arthritis, septic thrombophlebitis,
acute bacterial endocarditis, toxic shock syndrome, scalded skin
syndrome and Staphylococcal food poisoning. Epitopes of MurA
polypeptides and polynucleotides are useful immunogens for producing
anti-MurA antibodies for prevention of bacterial infections.

XX SQ Sequence 421 AA;

Query Match 42.9%; Score 6; DB 20; Length 421;


```

XX PF 10-DEC-1998; 98WO-AU01023.
XX PR 04-AUG-1998; 98AU-0005028.
XX PR 10-DEC-1997; 97AU-0000839.
XX PR 31-DEC-1997; 97AU-0001182.
XX PR 30-JAN-1998; 98AU-0001546.
XX PR 10-MAR-1998; 98AU-0002264.
XX PR 09-APR-1998; 98AU-0002911.
XX PR 23-APR-1998; 98AU-0003128.
XX PR 05-MAY-1998; 98AU-0003338.
XX PR 22-MAY-1998; 98AU-0003654.
XX PR 29-JUL-1998; 98AU-0004917.
XX PA (CSLC-) CSL LTD.
XX XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
XX PI Ross BC, Rothel LJ, Webb EA;
XX XX WPI; 1999-385613/32.
XX DR N-PSDB; AAX91726.
XX XX Antigenic Porphyromonas gingivalis peptides for preventing
XX PT gingivitis
XX XX Claim 1; Page 491-492; 588pp; English.
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphyromonas gingivalis. Probes can
XX CC be used to detect Porphyromonas gingivalis in standard hybridisation
XX CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX CC especially gingivitis.
XX SQ Sequence 867 AA;

Query Match 42.9%; Score 6; DB 20; Length 867;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RDTGCE 11
Db 522 rdytge 527

RESULT 11
AAY34381
XX ID AAY34381 standard; Protein: 875 AA.
XX AC AAY34381;
XX DT 25-AUG-1999 (first entry)
XX DE Porphyromonas gingivalis protein PG47.
XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX KW vaccine; antigenic.
XX OS Porphyromonas gingivalis.
XX OS WO9929870-A1.
XX PN 17-JUN-1999.
XX PD 10-DEC-1998; 98WO-AU01023.
XX PF 04-AUG-1998; 98AU-0005028.
XX PR 10-DEC-1997; 97AU-0000839.
XX PR 31-DEC-1997; 97AU-0001182.

```

```

XX PR 30-JAN-1998; 98AU-0001546.
XX PR 10-MAR-1998; 98AU-0002264.
XX PR 09-APR-1998; 98AU-0002911.
XX PR 23-APR-1998; 98AU-0003128.
XX PR 05-MAY-1998; 98AU-0003338.
XX PR 22-MAY-1998; 98AU-0003654.
XX PR 29-JUL-1998; 98AU-0004917.
XX PA (CSLC-) CSL LTD.
XX XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
XX PI Ross BC, Rothel LJ, Webb EA;
XX XX WPI; 1999-385613/32.
XX DR N-PSDB; AAX91599.
XX XX Antigenic Porphyromonas gingivalis peptides for preventing
XX PT gingivitis
XX XX Claim 1; Page 346-348; 588pp; English.
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphyromonas gingivalis. Probes can
XX CC be used to detect Porphyromonas gingivalis in standard hybridisation
XX CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX CC especially gingivitis.
XX SQ Sequence 875 AA;

Query Match 42.9%; Score 6; DB 20; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RDTGCE 11
Db 530 rdytge 535

RESULT 12
AAY01519
XX ID AAY01519 standard; Protein: 1462 AA.
XX AC AAY01519;
XX DT 14-JUN-1999 (first entry)
XX DE A carcinogenesis-inhibiting protein.
XX KW Carcinogenesis-inhibiting activity; genetic treatment; cerebral tumour.
XX KW Homo sapiens.
XX OS JPI1075844-A.
XX PN 23-MAR-1999.
XX PD 01-SEP-1997; 97JP-0236208.
XX PF 01-SEP-1997; 97JP-0236208.
XX PR (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX PA WPI; 1999-257694/22.
XX DR N-PSDB; X026546.
XX XX A carcinogenesis-inhibiting gene - useful for genetic treatment of
XX PT cerebral tumours
XX PT

```


PS Claim 1; Page 15-19; 21pp; Japanese.

XX The present sequence represents a protein with carcinogenesis-inhibiting

CC activity. The gene is useful for the genetic treatment of cerebral

CC tumours.

XX

SQ Sequence 1462 AA;

Query Match 42.9%; Score 6; DB 20; Length 1462;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 RDTGE 11

Db 1029 rdytge 1034

RESULT 13

AAU05396

ID AAU05396 standard; Protein; 26926 AA.

XX

AC AAU05396;

XX

DT 24-OCT-2001 (first entry)

XX

XX Human titin (connectin) protein sequence.

XX

XX Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;

KW titin-related disease; zebrafish; heart failure; heart disease.

XX

OS Homo sapiens.

XX

XX WO200151666-A1.

PN

XX

PD 19-JUL-2001.

XX

XX 12-JAN-2001; 2001WO-US01212.

XX

XX 12-JAN-2000; 2000US-0175787.

XX

XX (GEHO) GEN HOSPITAL CORP.

PA

XX

XX Fishman MC;

PI

XX

XX WPI: 2001-451869/48.

DR

DR N-PSDB; AAS05390.

XX

XX Determining if a subject has or is at risk of developing a

PT titin-related disease or condition, particularly heart failures,

PT comprises detecting the presence of a mutation in the titin gene

XX

PS Disclosure; Page 57-111; 114pp; English.

XX

CC The present sequence representing human titin (also known as connectin)

CC is described in an invention relating to a novel method for determining

CC whether a subject has or is at risk of developing a titin-related

CC disease or condition. The method comprises analysing a nucleic acid of

CC a sample from the subject and detecting the presence of a mutation

CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the

CC titin gene, which indicates that the subject has or is at risk of

CC developing a titin-related disease. The zebrafish which has a phenotype

CC similar to mammalian heart failure is used as a model. The method is

CC useful for detecting an increased likelihood of heart disease, such as

CC heart failure, in a patient, so that appropriate intervention can be

CC instituted before any symptoms occur. The method may also be used to

CC facilitate determination of etiology of an existing heart condition,

CC such as heart failure, to identify compounds that can be used to treat

CC or prevent heart conditions, in prenatal genetic screening, e.g. to

CC identify parents who may be carriers of a recessive titin mutation.

CC Compounds identified using the methods may be used to treat patients

CC that have or are at risk of developing heart disease, e.g. heart

CC failure.

XX

SQ Sequence 26926 AA;

Query Match 42.9%; Score 6; DB 22; Length 26926;

Best Local Similarity 100.0%; Pred. No. 2.8e+03;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 TGEVKV 14

Db 1861 tgevkv 1866

RESULT 14

AAW25728

ID AAW25728 standard; peptide; 9 AA.

XX

AC AAW25728;

XX

DT 20-APR-1998 (first entry)

XX

DE Labelled cholecystokinin analogue 2 (CKK 25-33) for detecting tumours.

XX

XX Cholecystokinin analogue; CKK receptor; detection; malignant;

KW localisation; tumour; magnetic resonance imaging; MRI; label.

XX

OS Synthetic.

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FH Modified-site 7 /label= Nle

FT Modified-site 9

FT Modified-site 9 /note= "C-terminal amide"

XX

XX WO9731657-A2.

PN

XX

XX 04-SEP-1997.

PD

XX

XX 25-FEB-1997; 97WO-US03056.

PF

XX

XX 27-FEB-1996; 96EP-0200498.

PR

XX

XX (MLCW) MALLINCKRODT MEDICAL INC.

PA

XX

XX Reubi J;

PI

XX

XX WPI: 1997-448445/41.

DR

XX

XX Use of labelled cholecystokinin analogues - for detecting,

PT localising or treating tumours and metastases in tissues which

PT normally do not contain CKK receptors

XX

PS Disclosure; Page 6; 61pp; English.

XX

CC The present sequence represents a cholecystokinin (CKK) analogue which

CC was used in a novel method for detecting and localising malignant

CC tumours and their metastases in tissues, which when healthy, do not

CC contain significant quantities of cholecystokinin (CKK) receptors. The

CC method comprises administering a peptide derived from the generic

CC formula (AAW25691). The peptide is labelled with a radioactive metal

CC isotope, a paramagnetic metal atom or a radioactive halogen isotope. The

CC human is subjected to external imaging by radioactive scanning or by

CC magnetic resonance imaging (MRI), to determine the targeted sites in the

CC body in relation to the background activity, in order to allow detection

CC and localisation of the tumours in the body.

XX

SQ Sequence 9 AA;

Query Match 35.7%; Score 5; DB 18; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.3e+05;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RDTG 10
Db 1 rdytg 5

Search completed: January 31, 2002, 13:18:14
Job time: 175 sec

RESULT 15

AAB37654
ID AAB37654 standard; peptide; 9 AA.

XX AC AAB37654;

XX XX 23-MAR-2001 (first entry)

XX DE Cholecystokinin peptide fragment analogue #33.

XX XX Antidiabetic; cytostatic; auditory; cholecystokinin-8; CCK-8; neuropathy;
KW peripheral nervous system; diabetes mellitus; cancer treatment;
KW cytotatica; hearing impairment; visual handicap;
KW alcohol-induced neuropathy; dystrophy.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 2 /note= "Asp(OBut)"

FT Modified-site 3 /note= "Tyr-(SO3Bal/2)"

FT Modified-site 4 /note= "Thr(But)"

FT Modified-site 8 /note= "Asp(OBut)"

FT Modified-site 9 /note= "C-terminal amide"

XX WO2000066150-A1.

XX PD 09-NOV-2000.

XX PF 03-MAY-2000; 2000WO-SE00870.

XX PR 03-MAY-1999; 99SE-0001578.

XX PA (KARO-) KAROLINSKA INNOVATIONS AB.

XX PI Lundeberg T, Manni L;

XX PS WPI; 2001-024739/03.

XX PT Manufacturing a medicament for treating neuropathies in the peripheral

XX PT nervous system comprises use of a substance showing cholecystokinin-8

XX PS activity

XX PS Disclosure; Page 6; 38pp; English.

XX CC The present invention relates to peptides showing cholecystokinin (CCK)-8

XX CC activity, which can be used to treat neuropathies in the peripheral

XX CC nervous system (PNS). The present sequence is one such peptide. The

XX CC peptides of the present invention may be used to treat neuropathies in

XX CC the PNS associated with diabetes mellitus, cancer treatment such as

XX CC cytotatica, hearing impairment and/or visual handicap, alcohol-induced

XX CC neuropathy, damage induced by surgery and dystrophy.

XX XX Sequence 9 AA;

Query Match 35.7%; Score 5; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 RDTG 10
Db 1 rdytg 5

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:40 ; Search time 130.99 seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-74
Perfect score: 14
Sequence: 1 AGWIDAGFKITL 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL17:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	50.0	564	5	Q23118 caenorhabdi
2	6	42.9	88	2	Q9K0J1 neisseria m
3	6	42.9	88	2	Q9J1V6 neisseria m
4	6	42.9	97	1	Q9HLC5
5	6	42.9	101	1	Q9P9E8
6	6	42.9	106	5	Q9NEV7
7	6	42.9	220	2	Q9ANU3
8	6	42.9	228	2	Q9F3E8
9	6	42.9	232	2	P95776 streptococ
10	6	42.9	233	1	P95931 sulfolobus
11	6	42.9	260	2	Q9L2D4 streptomyc
12	6	42.9	287	5	Q18483
13	6	42.9	296	10	Q9LZR5
14	6	42.9	311	2	O67636 aquifex ao
15	6	42.9	352	2	O66740
16	6	42.9	444	2	Q9X9K7
17	6	42.9	455	2	Q9JF33 streptomyc
18	6	42.9	465	5	O02206 neisseria m
19	6	42.9	519	2	Q99ZK1 caenorhabdi
					Q99ZK1 streptococ

20	6	42.9	525	10	Q9LZH4	Q9LZH4 arabidopsis
21	6	42.9	640	5	Q9GRE2	Q9GRF2 ptychodera
22	6	42.9	704	13	Q9I8G2	Q9I8G2 brachydanio
23	6	42.9	883	5	O76623	O76623 caenorhabdi
24	6	42.9	1011	5	Q9ND08	Q9NDQ8 ciona intes
25	6	42.9	1061	12	O41065	O41065 paramecium
26	6	42.9	1066	5	Q9NDQ9	Q9NDQ9 ciona intes
27	5	35.7	50	12	Q996F8	Q996F8 human echov
28	5	35.7	50	12	Q996F4	Q996F4 human echov
29	5	35.7	50	12	Q996F3	Q996F3 human echov
30	5	35.7	50	12	Q996F2	Q996F2 human echov
31	5	35.7	50	12	Q996F1	Q996F1 human echov
32	5	35.7	50	12	Q996F0	Q996F0 human echov
33	5	35.7	50	12	Q996E9	Q996E9 human echov
34	5	35.7	50	12	Q996E8	Q996E8 human echov
35	5	35.7	50	12	Q996E7	Q996E7 human echov
36	5	35.7	50	12	Q996E6	Q996E6 human echov
37	5	35.7	50	12	Q996E5	Q996E5 human echov
38	5	35.7	50	12	Q996E4	Q996E4 human echov
39	5	35.7	50	12	Q996E3	Q996E3 human echov
40	5	35.7	50	12	Q996E2	Q996E2 human echov
41	5	35.7	50	12	Q996D9	Q996D9 human echov
42	5	35.7	50	12	Q996D7	Q996D7 human echov
43	5	35.7	50	12	Q996D5	Q996D5 human echov
44	5	35.7	50	12	Q996D4	Q996D4 human echov
45	5	35.7	50	12	Q996D3	Q996D3 human echov

ALIGNMENTS

RESULT 1

Q23118 PRELIMINARY; PRT; 564 AA.

AC Q23118; 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

DT 01-JAN-1999 (TRENBLrel. 09, Last annotation update)

DE W01C9.2 PROTEIN.

GN W01C9.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OC NCBI_TaxID=6239;

EN [1]

RP SEQUENCE FROM N.A.

RA Sims M.;

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,

RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkison-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."

RL Nature 368:32-38(1994).

DR EMBL; Z49969; CAA90270.1; -.

SQ SEQUENCE 564 AA; 63998 MW; 1637EDDC4631D CRC64;

Query Match 50.0%; Score 7; DB 5; Length 564;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 FKGITL 14

```

|||||||
2 FRGKTL 8

RESULT 2
Q9K0J1 ID Q9K0J1 PRELIMINARY: PRT; 88 AA.
AC Q9K0J1 2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL PROTEIN NMB0606.
GN NMB0606.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cifton E., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Massignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RL Science 287:1809-1815(2000).
DR EMBL; AE024116; AAF41033.1;
DR TIGR; NMB0606;
DR InterPro; IPR003849; DUF219.
DR Pfam; PF02699; DUF219; 1.
DR Complete proteome.
SQ SEQUENCE 88 AA; 9709 MW; C751556C81A8EB44 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11
Db 50 AGFKGK 55
[1]

RESULT 3
Q9JVJ6 ID Q9JVJ6 PRELIMINARY: PRT; 88 AA.
AC Q9JVJ6 2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL PROTEIN NMA0811.
GN NMA0811
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222565; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis J., Devlin K., Feitwell T., Hamlin N., Holroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whithead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RL meningitidis Z2491."
RL Nature 404:502-506(2000).

```

```

DR EMBL; AL162754; CAB84093.1;
DR InterPro; IPR003849; DUF219.
DR Pfam; PF02699; DUF219; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 88 AA; 9725 MW; C751556C9C8EE44 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11
Db 50 AGFKGK 55
[1]

RESULT 4
Q9HLC5 ID Q9HLC5 PRELIMINARY: PRT; 97 AA.
AC Q9HLC5 2000 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE HYPOTHETICAL PROTEIN TAO304.
GN TAO304.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RL Nature 407:508-513(2000).
DR EMBL; AL445063; CAC11449.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 97 AA; 11207 MW; DEC76C6412855FDD CRC64;

Query Match 42.9%; Score 6; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GWIDAG 7
Db 73 GWIDAG 78
[1]

RESULT 5
Q9P9E8 ID Q9P9E8 PRELIMINARY: PRT; 101 AA.
AC Q9P9E8 2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL 11.4 KDA PROTEIN.
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JJ;
RA Lin W.C.; Whitman W.B.;
RT "Cloning of Methanococcus maripaludis pyruvate oxidoreductase."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF230199; AAF91261.1;
KW Hypothetical protein.
SQ SEQUENCE 101 AA; 11426 MW; BD7A24CFDF64152F CRC64;

```

```

Query Match          42.9%; Score 6; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GWIDAG 7
    |||||
Db 77 GWIDAG 82

RESULT 6
Q9NEY7
ID Q9NEY7 PRELIMINARY; PRT; 106 AA.
AC Q9NEY7;
DT 01-OCT-2000 (TREMELREL. 15, Created)
DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE Y1058E.M PROTEIN.
GN Y1058E.M.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL132880; CAB60876.2; -.
DR InterPro; IPR002634; BOLA.
DR Pfam; PF01722; BOLA; 1.
SQ SEQUENCE 106 AA; 11423 MW; 2B969AB2D590D360 CRC64;

Query Match          42.9%; Score 6; DB 5; Length 106;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11
    |||||
Db 71 AGFKGK 76

RESULT 7
Q9ANU3
ID Q9ANU3 PRELIMINARY; PRT; 220 AA.
AC Q9ANU3;
DT 01-JUN-2001 (TREMELREL. 17, Created)
DT 01-JUN-2001 (TREMELREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)
DE 1.2-DIHYDROXYNAPHTHALENE DIOXYGENASE 2 (FRAGMENT).
GN NAHC2.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SIITASAL;
RA Ferrero M.A., Llobet-Brossa E., Lallucat J., Rossello-Mora R.,
RA Bosch R.;
RT "Coexistence of two distinct naphthalene degradation upper pathways in
RT Pseudomonas isolated from the West Mediterranean Region."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF320641; AAG53395.1; -.
KW DIOXYGENASE.

Query Match          42.9%; Score 6; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GWIDAG 7
    |||||
Db 182 GWIDAG 187

RESULT 9
P95776
ID P95776 PRELIMINARY; PRT; 232 AA.
AC P95776;
DT 01-MAY-1997 (TREMELREL. 03, Created)
DT 01-MAY-1997 (TREMELREL. 03, Last sequence update)
DT 01-JAN-1999 (TREMELREL. 09, Last annotation update)
DE ORF3 PROTEIN.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-XC;
RX MEDLINE=971175539; PubMed=9023194;
RA Tsukitaka Y., Yamashita Y., Oho T., Nakano Y., Koga T.;
RT "Biological function of the TDP-rhamnose synthesis pathway in
RL Streptococcus mutans.";
RQ EMBL; D78182; BAAL1245.1; -.
SQ SEQUENCE 232 AA; 26030 MW; D078E11994604525 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KGKITL 14
   |||||
Db 165 KGKITL 170

RESULT 10
P95931 PRELIMINARY; PRT; 233 AA.
AC P95931;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ORF C01033.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1617 / P2;
RX MEDLINE=97055432; PubMed=8899719;
RA Sensen C.W., Klenk H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
RA Liu Q.Y., Penny S.B., Young F., Schenk M.E., Gaasterland T.,
RA Doolittle W.F., Regan M.A., Charlebois R.L.;
RT "Organizational characteristics and information content of an archaeal
RT genome: 156 kb of sequence from Sulfolobus solfataricus P2.";
RL Mol. Microbiol. 22:1175-191(1996).
DR EMBL; Y08256; CAA69435.1; -.
SQ SEQUENCE 233 AA; 26215 MW; 5E3AE64515CAAD90 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFKGKI 12
   |||||
Db 164 GFKGKI 169

RESULT 11
Q9L2D4 PRELIMINARY; PRT; 260 AA.
AC Q9L2D4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 26.0 KDA PROTEIN.
GN SC7A8.24C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RQ EMBL; D78182; BAAL1245.1; -.
SQ SEQUENCE FROM N.A.

```

```

RC STRAIN-A3(2);
RA Cordero A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN 31
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL137187; CAB69773.1; -.
DR InterPro; IPR002781; DUF81.
DR Pfam; PF01925; DUF81; 2.
KW Hypothetical protein.
SQ SEQUENCE 260 AA; 25996 MW; DA0F43E1197BACE1 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGWIDA 6
   |||||
Db 19 AGWIDA 24

RESULT 12
Y18483 PRELIMINARY; PRT; 287 AA.
ID Q18483
AC Q18483;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SIMILAR TO ALDOSE REDUCTASES.
GN C35D10.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Ropra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Fulton L.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN 131
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U21324; AAA62562.1; -.
DR HSSP; P80276; IAH4.
DR InterPro; IPR001395; Aldo_ket_red.
DR Pfam; PF00248; aldo_ket_red; 1.
RQ PRINTS; PR00069; ALDKETRDASE.

```

DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; UNKNOWN_1.
SQ SEQUENCE 287 AA; 32840 MW; 9E9B49795E2D054 CRC64;

Query Match 42.9%; Score 6; DB 5; Length 287;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 IDAGFK 9
|||||

DB 31 IDAGFK 36

RESULT 13

Q9LZR5 09LZR5 PRELIMINARY; PRT; 296 AA.

AC 09LZR5

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE HISTONE DEACETYLASE-LIKE PROTEIN.

GN F17C15_160.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W., Rudd S.,

RA Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL162506; CAB82939.1; -

DR InterPro; IPR000822; Znf-C2H2.

DR Pfam; PF00096; zf-C2H2; 1.

DR SMART; SM00355; Znf-C2H2; 1.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.

DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.

KW DNA-binding; Metal-binding; Zinc-finger.

SQ SEQUENCE 296 AA; 32129 MW; 4B2DAABEAALF1E9D CRC64;

Query Match 42.9%; Score 6; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11
|||||

DB 105 AGFKGK 110

RESULT 14

O67636 067636 PRELIMINARY; PRT; 311 AA.

AC 067636

DT 01-AUG-1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE PREPHENATE DEHYDROGENASE.

GN TYRA OR AQ_1755.

OS Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.

OX NCBI_TaxID=63363;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VF5;

RX MEDLINE-98196666; PubMed-9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,

RA Graham D.E., Overbeek R., Snead M.A., AuJay M., Huber R.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";

RL Nature 392:353-358(1998).

DR EMBL; AE000754; AAC07589.1; -

DR InterPro; IPR000205; NAD_binding.

DR InterPro; IPR003099; PDH.

DR InterPro; IPR000594; ThIF_family.

DR Pfam; PF02153; PDH; 1.

KW Complete proteome.

SQ SEQUENCE 311 AA; 34849 MW; CCA6DA3EFC7A18CD CRC64;

Query Match 42.9%; Score 6; DB 2; Length 311;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFKGKI 12
|||||

DB 53 GFKGKI 58

RESULT 15

O66740 066740 PRELIMINARY; PRT; 352 AA.

AC 066740

DT 01-AUG-1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE THREONINE SYNTHASE.

GN THRC2 OR AQ_425.

OS Aquifex aeolicus.

OC Bacteria; Aquificales; Aquificaceae; Aquifex.

OX NCBI_TaxID=63363;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-VF5;

RX MEDLINE-98196666; PubMed-9537320;

RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,

RA Graham D.E., Overbeek R., Snead M.A., Kellier M., AuJay M., Huber R.,

RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;

RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";

RL Nature 392:353-358(1998).

DR EMBL; AE000689; AAC06690.1; -

DR InterPro; IPR000634; dehydrtse_ser_thr.

DR InterPro; IPR001926; PALP.

DR Pfam; PF00291; PALP; 1.

DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.

KW Complete proteome.

SQ SEQUENCE 352 AA; 37808 MW; 3BF480E6E097ACD2 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 352;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFKGKI 12
|||||

DB 44 GFKGKI 49

Search completed: January 31, 2002, 13:37:42
Job time: 167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:18 ; Search time 46.78 Seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-74
Perfect score: 14
Sequence: 1 AGWIDAGFKGKITL 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	14	100.0	204	1 DCD_METJA	Q57872 methanococ
2	7	50.0	315	1 TRAB_EUBAC	P50971 eubacterium
3	7	50.0	443	1 YEY2_YEAST	P10356 saccharomyc
4	6	42.9	103	1 Y273_METJA	Q57721 methanococ
5	6	42.9	183	1 Y389_HAEIN	P43991 haemophilus
6	6	42.9	323	1 MOAA_CLOPE	Q9wx96 clostridium
7	6	42.9	329	1 PSBO_PEA	P14226 pisum sativ
8	6	42.9	338	1 ETEA_MEGEL	O85892 megasphaera
9	6	42.9	356	1 HITC_HAEIN	P44513 haemophilus
10	6	42.9	508	1 GSBP_CHICK	P12244 gallus gall
11	6	42.9	509	1 PDI_MOUSE	P09103 mus musculu
12	6	42.9	509	1 PDI_RABIT	P21195 oryctolagus
13	6	42.9	509	1 PDI_RAT	P04785 rattus norv
14	6	42.9	575	1 YDGE_SCHPO	Q10499 schizosacch
15	6	42.9	731	1 MASZ_MYCLE	O32913 mycobacteri
16	6	42.9	1132	1 VHSJ_LAMBD	P03749 bacterioph
17	6	42.9	1160	1 TFC3_YEAST	P34111 saccharomyc
18	6	42.9	1785	1 BIC2_HUMAN	Q9V6D5 homo sapien
19	6	42.9	1849	1 BIG1_BOVIN	O46382 bos taurus
20	6	42.9	1849	1 BIG1_HUMAN	Q9Y6D6 homo sapien
21	5	35.7	47	1 SA21_SORBI	O09198 sorghum bic
22	5	35.7	47	1 SA13_SORBI	P21925 sorghum bic
23	5	35.7	47	1 THG1_WAIZE	P81008 zea mays (m
24	5	35.7	47	1 THG1_WHEAT	P20158 triticum ae
25	5	35.7	47	1 THG2_WHEAT	P20159 triticum ae
26	5	35.7	47	1 THG_HORVU	P20230 hordeum vul
27	5	35.7	48	1 SA12_SORBI	P21924 sorghum bic
28	5	35.7	91	1 RK15_SPIOL	P22798 spinacia ol
29	5	35.7	92	1 ABH_BACSU	P39758 bacillus su
30	5	35.7	115	1 IFJA_PVRAB	Q9V138 pyrococcus
31	5	35.7	116	1 IFJA_PYRHO	O59280 pyrococcus
32	5	35.7	119	1 P15_RAT	Q63396 rattus norv
33	5	35.7	126	1 P15_MOUSE	P11031 mus musculu

RESULT 1
DCD_METJA ID DCD_METJA STANDARD; PRT; 204 AA.
AC Q57872;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE (EC 3.5.4.13) (DCTP
DE DEAMINASE).
GN DCD OR M30430.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
Science 273:1058-1073(1996).
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: DCTP + H(2)O -> DUTP + NH(3).
CC -!- SIMILARITY: BELONGS TO THE DCTP DEAMINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67494; AAB98415.1; -
CC TIGR: M30430;
DR InterPro: IPR003232; dCTP_deaminase.
DR InterPro: IPR001428; dUTPase.
DR Pfam: PF006592; dUTPase; 1.
DR ProDom: PD004900; dCTP_deaminase; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 204 AA; 23432 MW; 1218368057723371 CRC64;

Query Match 100.0%; Score 14; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 6.9e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGWIDAGFKGKITL 14
|||||

AC Q37721;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL PROTEIN MJ0273.
GN MJ0273.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.;
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67482; AAB98261.1; -;
DR TIGR; MJ0273; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11972 MW; BDE81755780AF7EB CRC64;

Query Match 42.9%; Score 6; DB 1; Length 103;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GWIDAG 7
Db 79 GWIDAG 84
| | | | |

RESULT 5
Y389_HAEIN
ID Y389_HAEIN STANDARD; PRT; 183 AA.
AC P43991;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOHETICAL PROTEIN HI0389.
GN HI0389.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagan N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32722; AAC22047.1; -;
DR TIGR; HI0389; -;
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSEM 7 23 POTENTIAL.
SQ SEQUENCE 183 AA; 21170 MW; 602C303C10A9B61B CRC64;

Query Match 42.9%; Score 6; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 KGKITL 14
Db 2 KGKITL 7
| | | | |

RESULT 6
MOAA_CLOPE
ID MOAA_CLOPE STANDARD; PRT; 323 AA.
AC Q9WX96;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN A.
GN MOAA.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB56K;
RA Okabe A.;
RT *Clostridium perfringens nitrate reductase.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF MOLYBDOPTERIN PRECURSOR
CC Z FROM GUANOSINE (BY SIMILARITY).
CC -!- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE MOAA / NIFB / PQQE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB017192; BAA76928.1; -;
DR InterPro; IPR000385; Moaa_NifB_Pqqe.
DR Pfam; PF01444; Moaa_NifB_Pqqe; 1.
DR PROSITE; PS01305; MOAA_NIFB_PQQE; 1.
KW Molybdenum cofactor biosynthesis; Iron-sulfur.
FT METAL 20 20 IRON-SULFUR (POTENTIAL).
FT METAL 24 24 IRON-SULFUR (POTENTIAL).
FT METAL 27 27 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 323 AA; 37390 MW; 1BA50E90B850FEC0 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 323;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFGKGI 12
Db 234 GFGKGI 239
|||||

RESULT 7
PSBO_PEA 7
ID PSBO_PEA STANDARD; PRT; 329 AA.
AC P14226;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OXYGEN-EVOLVING ENHANCER PROTEIN 1, CHLOROPLAST PRECURSOR (OEEL1) (33
DE KDA SUBUNIT OF OXYGEN EVOLVING SYSTEM OF PHOTOSYSTEM II) (OEC 33 KDA
DE SUBUNIT) (33 KDA THYLAKOID MEMBRANE PROTEIN).
GN PSBO.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC Eurosid 1; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RA Wiles R., Newman B.J., Pappin D., Gray J.C.;
RA "The extrinsic 33 kDa polypeptide of the oxygen-evolving complex of
RA photosystem II is a putative calcium-binding protein and is encoded by
RA multi-gene family in pea."
RA Plant Mol. Biol. 12:439-451(1989).
RP SEQUENCE FROM N.A.
RA Matanabe A., Shinohara K., Murase M., Maruta Y., Konishi T.;
RA Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: STABILIZES THE MANGANESE CLUSTER WHICH IS THE PRIMARY
CC SITE OF WATER SPLITTING (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE; ASSOCIATED
CC WITH THE PHOTOSYSTEM II COMPLEX.
CC -1- SIMILARITY: BELONGS TO THE PSBO FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X15350; CAA33408.1; -
CC EMBL: D13297; BAA02554.1; -
CC PIR: S04132; S04132.
CC Mendel; 16708; PISA: Psbo; mml6708.
CC InterPro: IPR002628; MSP.
CC Pfam: PF01716; MSP; 1.
CC Photosynthesis; Photosystem II; Chloroplast; Transit peptide;
CC Thylakoid; Membrane; 81
FT TRANSIT 1
FT CHAIN 82 329 OXYGEN-EVOLVING ENHANCER PROTEIN 1.
SQ SEQUENCE 329 AA; 34893 MW; 8F5F787616C5D8E6 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 329;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GKITL 14
Db 278 GKITL 283
|||||

RESULT 8

ETFA_MEGEL 42.9%; Score 6; DB 1; Length 338 AA.
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGWIDA 6
Db 255 AGWIDA 260
|||||

RESULT 9
HITC_HAEIN 9
ID HITC_HAEIN STANDARD; PRT; 356 AA.
AC P44513; Q53441;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE IRON(III)-TRANSFER ATP-BINDING PROTEIN HITC.
DE HITC OR H10099.
GN Haemophilus influenzae.
OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-RD / KW20 / ATCC 51907;
RC MEDLINE-95350630; PubMed-7542800;
RX

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RT Science 269:496-512(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI TN106;
RX MEDLINE=95012644; PubMed=7927717;
RA Sanders J.D., Cope L.D., Hansen E.J.;
RT "Identification of a locus involved in the utilization of iron by
RT Haemophilus influenzae.";
RL Infect. Immun. 62:4515-4525(1994).
CC -!- FUNCTION: NUCLEOTIDE-BINDING PROTEIN COMPONENT OF THE PERIPLASMIC
CC BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR FE(III).
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32695; AAC21775.1; -;
DR EMBL: S72674; BAB32112.1; -;
DR HSSP: P13569; INBD.
DR TIGR: H10099; -;
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00005; ABC_tran; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
DR Iron transport; Inner membrane; Transport; ATP-binding;
KW Complete proteome.
FT NP_BIND 44 51 ATP (POTENTIAL).
FT VARIANT 45 45 S -> A (IN STRAIN TN106).
FT VARIANT 97 97 V -> I (IN STRAIN TN106).
FT VARIANT 209 209 S -> A (IN STRAIN TN106).
FT VARIANT 350 350 R -> K (IN STRAIN TN106).
FT VARIANT 356 356 S -> A (IN STRAIN TN106).
FT SEQUENCE 356 AA; 40360 MW; 78F862F359760813 CRC64;
SQ
Query Match 42.9%; Score 6; DB 1; Length 356;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 FKGIIT 13
Db 312 FKGIIT 317
RESULT 10
GSBP_CHICK STANDARD; PRT; 508 AA.
AC P12244;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE DOLICHYL-DIPHOSPHOOLIGOSACCHARIDE-PROTEIN GLYCOTRANSFERASE PRECURSOR
DE (EC 2.4.1.119) (GLYCOSYLATION SITE-BINDING CHAIN) (GSBP).

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88327849; PubMed=2458190;
RA Geetha-Habib M., Noiva R., Kaplan H.A., Lennarz W.J.;
RT "Glycosylation site binding protein, a component of oligosaccharyl
RT transferase, is highly similar to three other 57 kd luminal proteins
RT of the ER.";
RL Cell 54:1053-1060(1988).
CC -!- FUNCTION: TRANSFER THE POLYSACCHARIDE SIDE-CHAINS OF GLYCOPROTEINS
CC TO AN ASPARAGINE RESIDUE OF ASN-X-SER/THR SITES IN NASCENT
CC PROTEINS.
CC -!- CATALYTIC ACTIVITY: DOLICHYL DIPHOSPHOOLIGOSACCHARIDE + PROTEIN
CC L-ASPARAGINE = DOLICHYL DIPHOSPHATE + A GLYCOPROTEIN WITH THE
CC OLIGOSACCHARIDE CHAIN ATTACHED BY GLYCOSAMINE LINKAGE TO PROTEIN
CC L-ASPARAGINE.
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -!- SIMILARITY: TO PROTEIN DISULFIDE ISOMERASE.
CC -!- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M22594; AAA64295.1; -;
DR PIR: A30007; A30007.
DR HSSP: P07237; IMEX.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR000063; Thioered.
DR Pfam: PF00085; Thioered; 2.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00194; THIOREDOXIN; 1.
KW Redox-active center; isomerase; Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 508 GSBP.
FT DISULFID 52 55 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 398 401 REDOX-ACTIVE (BY SIMILARITY).
FT SITE 505 508 PREVENT SECRETION FROM ER.
FT SEQUENCE 508 AA; 56892 MW; D2BA0E5872BE58BE CRC64;
SQ
Query Match 42.9%; Score 6; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 GFKGKI 12
Db 281 GFKGKI 286
RESULT 11
PDI_MOUSE STANDARD; PRT; 509 AA.
AC P09103;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) (EC 5.3.4.1) (PROLYL 4-
DE HYDROXYLASE BETA SUBUNIT) (CELLULAR THYROID HORMONE BINDING PROTEIN
DE (P55) (ERP59).
GN PDI4L OR P4HB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88143996; PubMed=2830592;
RA Gong Q.-H., Fukuda T., Parkinson C., Cheng S.-Y.;
RT "Nucleotide sequence of a full-length cDNA clone encoding a mouse
RT cellular thyroid hormone binding protein (p55) that is homologous to
RT protein disulfide isomerase and the beta-subunit of prollyl-4-
RT hydroxylase.";
RL Nucleic Acids Res. 16:1203-1203(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90110091; PubMed=2295602;
RA Mazzarella R.A., Srinivasan M., Haugetjorden S.M., Green M.;
RT "Prp72, an abundant luminal endoplasmic reticulum protein, contains
RT three copies of the active site sequences of protein disulfide
RT isomerase.";
RL J. Biol. Chem. 265:11094-1101(1990).
CC -1- FUNCTION: PDI, THE BETA SUBUNIT OF PROLYL 4-HYDROXYLASE, AND THE
CC CELLULAR THYROID HORMONE BINDING PROTEIN ARE IDENTICAL.
CC -1- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN
CC DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
CC -1- CATALYTIC ACTIVITY: PROCOLLAGEN L-PROLINE + 2-OXOGLUTARATE + O(2)
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X06453; CAA29759.1;
DR EMBL: J05185; AAA39906.1;
DR PIR: A34930; ISWSS.
DR HSP: P07237; IMEX.
DR SWISS-2DPAGE: P09103; MOUSE.
DR MGI: 97464; P4Hb.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR000063; Thioired.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00194; THIOREDOXIN; 2.
KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 19
FT CHAIN 20 509
FT DISULFID 55 58
FT DISULFID 399 402
FT SITE 506 509
FT CONFLICT 68 68
FT SEQUENCE 509 AA; 57143 MW; DB6B3F5851088731 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 509;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFGKGI 12
Db 283 GFGKGI 288

RESULT 12
PDI_RABIT STANDARD; PRT; 509 AA.
AC P21195;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Last sequence update)

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88143996; PubMed=2830592;
RA Gong Q.-H., Fukuda T., Parkinson C., Cheng S.-Y.;
RT "Nucleotide sequence of a full-length cDNA clone encoding a mouse
RT cellular thyroid hormone binding protein (p55) that is homologous to
RT protein disulfide isomerase and the beta-subunit of prollyl-4-
RT hydroxylase.";
RL Nucleic Acids Res. 16:1203-1203(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90110091; PubMed=2295602;
RA Mazzarella R.A., Srinivasan M., Haugetjorden S.M., Green M.;
RT "Prp72, an abundant luminal endoplasmic reticulum protein, contains
RT three copies of the active site sequences of protein disulfide
RT isomerase.";
RL J. Biol. Chem. 265:11094-1101(1990).
CC -1- FUNCTION: PDI, THE BETA SUBUNIT OF PROLYL 4-HYDROXYLASE, AND THE
CC CELLULAR THYROID HORMONE BINDING PROTEIN ARE IDENTICAL.
CC -1- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN
CC DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
CC -1- CATALYTIC ACTIVITY: PROCOLLAGEN L-PROLINE + 2-OXOGLUTARATE + O(2)
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X06453; CAA29759.1;
DR EMBL: J05185; AAA39906.1;
DR PIR: A34930; ISWSS.
DR HSP: P07237; IMEX.
DR SWISS-2DPAGE: P09103; MOUSE.
DR MGI: 97464; P4Hb.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR000063; Thioired.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00194; THIOREDOXIN; 2.
KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 19
FT CHAIN 20 509
FT DISULFID 55 58
FT DISULFID 399 402
FT SITE 506 509
FT CONFLICT 68 68
FT SEQUENCE 509 AA; 57143 MW; DB6B3F5851088731 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 509;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFGKGI 12
Db 283 GFGKGI 288

RESULT 13
PDI_RAT STANDARD; PRT; 509 AA.
AC P04785; P13700;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) (EC 5.3.4.1) (PROLYL 4-
DE HYDROXYLASE BETA SUBUNIT) (CELLULAR THYROID HORMONE BINDING PROTEIN)
DE (THYROXINE DEIODINASE) (EC 3.8.1.4) (IODOETHYRONE 5'-MONODEIODINASE)
DE (5'-MD).
GN PDIAL OR P4HB.
OS Rattus norvegicus (Rat).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=86014354; PubMed=3840230;
 RA Edman J.C., Ellis L., Blacher R.W., Roth R.A., Rutter W.J.;
 RT "Sequence of protein disulphide isomerase and implications of its
 relationship to thioredoxin.";
 RL Nature 317:267-270(1985).
 RN [2]
 RP SEQUENCE OF 28-509 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=89025800; PubMed=3178809;
 RA Boado R.J., Campbell D.A., Chopra I.J.;
 RT "Nucleotide sequence of rat liver iodothyronine 5'-monodeiodinase (5'
 MD): its identity with the protein disulfide isomerase.";
 RL Biochem. Biophys. Res. Commun. 155:1297-1304(1988).
 RN [3]
 RP SEQUENCE OF 129-394 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=88296303; PubMed=2841089;
 RA Boado R.J., Chopra I.J., Flink I.L., Campbell D.A.;
 RT "Enzyme binding-inhibiting assay for iodothyronine 5'-monodeiodinase
 (5'-MD) and its application to isolation of complementary
 deoxyribonucleic acid clones for the 5'-MD in rat liver.";
 RL Endocrinology 123:1264-1273(1988).
 RN [4]
 RP SEQUENCE OF 20-34.
 RC STRAIN=LEC; TISSUE=Liver;
 RX MEDLINE=94072621; PubMed=8251535;
 RA Yokoi T., Nagayama S., Kajiwara R., Kawaguchi Y., Horiuchi R.,
 RA Kametaki T.;
 RT "Identification of protein disulfide isomerase and calreticulin as
 autoantigen antigens in LEC strain of rats.";
 RL Biochim. Biophys. Acta 1158:339-344(1993).
 CC -1- FUNCTION: PDI, THE BETA SUBUNIT OF PROLYL 4-HYDROXYLASE, AND THE
 CELLULAR THYROID HORMONE BINDING PROTEIN ARE IDENTICAL.
 CC -1- CATALYTIC ACTIVITY: REARRANGEMENT OF BOTH INTRACHAIN & INTERCHAIN
 DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES.
 CC -1- CATALYTIC ACTIVITY: PROCOLLAGEN L-PROLINE + 2-OXOGLUTARATE + O(2)
 = PROCOLLAGEN TRANS-4-HYDROXY-L-PROLINE + SUCCINATE + CO(2).
 CC -1- CATALYTIC ACTIVITY: L-THYRONINE + ALH(2) = 3,5,3'-L-THIODO-L-
 THYRONINE + IODIDE + A + H(+).
 CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; M21018; AAA40620.1; -
 DR EMBL; X02918; CAA26675.1; -
 DR EMBL; M21476; AAA40619.1; -
 DR PIR; A24595; ISRTSS.
 DR PIR; A31118; A31118.
 DR PIR; S06419; S06419.
 DR HSSP; P07237; IMEX.
 DR InterPro: IPR000886; ER.target.
 DR InterPro: IPR000083; ThioRed.
 DR Pfam; PF00085; ThioRed; 2.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR PROSITE; PS00014; ER_TARGET; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 2.
 KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal;
 HYDROLASE.
 FT SIGNAL 1 19
 FT CHAIN 20 509 PROTEIN DISULFIDE ISOMERASE.

FT REPEAT 28 110
 FT REPEAT 373 451
 FT REPEAT 276 363
 FT REPEAT 173 264
 FT DISULFID 55 58
 FT DISULFID 399 402
 FT SITE 506 509
 FT CONFLICT 39 40
 SQ SEQUENCE 509 AA; 56951 MW; 3056107F5E8B1B54 CRC64;
 Query Match 42.9%; Score 6; DB 1; Length 509;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 GFKGKI 12
 DB 283 GFKGKI 288
 RESULT 14
 YDGE_SCHPO STANDARD; PRT; 575 AA.
 ID YDGE_SCHPO
 AC Q10499; O14004;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PUTATIVE FLAVOPROTEIN C28F1.14C.
 GN SPAC26F1.14C OR SPAC29A4.01C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE OF 1-256 FROM N.A.
 RC STRAIN=972;
 RA Brown D., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 222-575 FROM N.A.
 RC STRAIN=972;
 RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; Z97210; CAB55362.2; -
 DR EMBL; Z73100; CAA97371.1; -
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR InterPro: IPR001281; Rieske.
 DR Pfam; PF00070; pyr_redox; 1.
 DR Pfam; PF00355; Rieske; 1.
 DR PRINTS; PR00368; FADPNR.
 DR KW Hypothetical protein; Flavoprotein.
 SQ SEQUENCE 575 AA; 62102 MW; 82F88DD1E12E638 CRC64;
 Query Match 42.9%; Score 6; DB 1; Length 575;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 FKGGIT 13
 DB 181 FKGGIT 186

RESULT 15
MASZ_MYCLE
ID MASZ_MYCLE STANDARD; PRT; 731 AA.
AC O32913;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE MALATE SYNTHASE G (EC 4.1.3.2).
GN GLCB OR ML2069 OR MLCB1788.27.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RA MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Bigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus".
RL Nature 409:1007-1011(2001).
CC -1- CATALYTIC ACTIVITY: L-MALATE + COA = ACETYL-COA + H(2)O +
CC GLYOXYLATE
CC -1- PATHWAY: SECOND STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE
CC TRICARBOXYLIC ACID CYCLE (IN BACTERIA, FUNGI AND PLANTS).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MALATE SYNTHASE G FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL008609; CAA15459.1; -;
CC EMBL: AL583924; CAC31024.1; -;
CC Leproma; ML2069; -;
KW Glyoxylate bypass; Tricarboxylic acid cycle; Lyase; Complete proteome.
SQ SEQUENCE 731 AA; 80141 MW; 3878CADA45DB416C CRC64;

Query Match 42.9%; Score 6; DB 1; Length 731;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11
| | | | |
Db 508 AGFKGK 513

Search completed: January 31, 2002, 13:39:19
Job time: 79 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:11 ; Search time 65.13 seconds
(without alignments)
4.837 Million cell updates/sec

Title: US-08-957-709-74
Perfect score: 14
Sequence: 1 AGWIDAGFKGKITL 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	4	US-08-822-774-55
2	6	42.9	508	2	US-08-557-122A-37
3	6	42.9	509	2	US-08-557-122A-27
4	6	42.9	510	1	US-08-441-139-20
5	6	42.9	510	2	US-08-557-122A-30
6	6	42.9	3052	2	US-08-557-122A-26
7	5	35.7	7	4	US-09-296-284-13
8	5	35.7	10	3	US-08-159-339A-833
9	5	35.7	13	5	PCT-US96-00206-2
10	5	35.7	13	5	PCT-US96-00206-3
11	5	35.7	16	5	PCT-US96-00206-4
12	5	35.7	22	2	US-08-521-871A-5
13	5	35.7	26	1	US-07-951-565-1
14	5	35.7	26	1	US-07-951-565-6
15	5	35.7	26	1	US-08-246-242-9
16	5	35.7	26	5	PCT-US96-00206-1
17	5	35.7	47	1	US-08-377-687-30
18	5	35.7	47	1	US-08-556-318A-12
19	5	35.7	47	2	US-08-777-192-30
20	5	35.7	47	2	US-08-956-459-12
21	5	35.7	47	4	US-08-971-982-30
22	5	35.7	48	1	US-08-377-687-29
23	5	35.7	48	1	US-08-556-318A-11
24	5	35.7	48	2	US-08-777-192-29
25	5	35.7	48	2	US-08-956-459-11
26	5	35.7	48	4	US-08-971-982-29
27	5	35.7	53	2	US-08-316-650-10

28	5	35.7	53	3	US-08-479-722B-10	Sequence 10, Appl
29	5	35.7	53	5	PCT-US95-02251-10	Sequence 10, Appl
30	5	35.7	100	3	US-08-946-026-21	Sequence 21, Appl
31	5	35.7	116	4	US-09-370-473-10	Sequence 10, Appl
32	5	35.7	124	3	US-09-187-049-8	Sequence 8, Appl
33	5	35.7	125	3	US-09-187-049-7	Sequence 7, Appl
34	5	35.7	213	3	US-08-523-373-22	Sequence 22, Appl
35	5	35.7	214	3	US-08-523-373-23	Sequence 23, Appl
36	5	35.7	215	3	US-08-523-373-24	Sequence 24, Appl
37	5	35.7	235	3	US-09-167-434-6	Sequence 6, Appl
38	5	35.7	235	3	US-08-853-755-6	Sequence 6, Appl
39	5	35.7	279	4	US-09-010-999-2	Sequence 2, Appl
40	5	35.7	280	4	US-09-296-284-27	Sequence 27, Appl
41	5	35.7	306	4	US-09-296-284-6	Sequence 6, Appl
42	5	35.7	341	2	US-08-846-762-92	Sequence 92, Appl
43	5	35.7	344	1	US-08-657-192-3	Sequence 3, Appl
44	5	35.7	344	3	US-08-523-373-5	Sequence 5, Appl
45	5	35.7	363	1	US-08-530-950-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-822-774-55
; Sequence 55, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-08-822-774-55

Query Match 100.0%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 6.6e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGWIDAGFKGKITL 14

```
Db      1  |||||AGWIDAGFKGKIL 14

RESULT 2
US-08-557-122A-37
; Sequence 37, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-557-122A-37

Query Match 42.9%; Score 6; DB 2; Length 508;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7  GFKGKI 12
Db      282 GFKGKI 287

RESULT 3
US-08-557-122A-27
; Sequence 27, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 06-JUL-1993
; APPLICANT INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

Query Match 42.9%; Score 6; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7  GFKGKI 12
Db      283 GFKGKI 288

RESULT 4
US-08-441-139-20
; Sequence 20, Application US/08441139
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Witttrup, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; APPLICANT INFORMATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
US-08-441-139-20

Query Match          42.9%; Score 6; DB 1; Length 510;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GFKGKI 12
Db 282 GFKGKI 287

RESULT 5
US-08-557-122A-30
; Sequence 30, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5879664 disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-557-122A-30

Query Match          42.9%; Score 6; DB 2; Length 510;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GFKGKI 12
Db 282 GFKGKI 287

RESULT 6
US-08-557-122A-26
; Sequence 26, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5879664 disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 510 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-557-122A-30

Query Match          42.9%; Score 6; DB 2; Length 510;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GFKGKI 12
Db 282 GFKGKI 287

RESULT 7
US-09-296-284-13
; Sequence 13, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-13

Query Match          35.7%; Score 5; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DAGFK 9
Db 3 DAGFK 7

RESULT 8
US-08-159-339A-833
; Sequence 8, Application US/08159339A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5879664 disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3052 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-557-122A-26

Query Match          42.9%; Score 6; DB 2; Length 3052;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GFKGKI 12
Db 1462 GFKGKI 1467

RESULT 7
US-09-296-284-13
; Sequence 13, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Gluconobacter suboxydans
US-09-296-284-13

Query Match          35.7%; Score 5; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 DAGFK 9
Db 3 DAGFK 7

RESULT 8
US-08-159-339A-833
; Sequence 8, Application US/08159339A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5879664 disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3052 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-557-122A-26
```

; Sequence 833, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Cellis, Esteban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,339A
; FILING DATE: 29-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/926,666
; FILING DATE: 07-AUG-1992
; APPLICATION NUMBER: US 08/027,746
; FILING DATE: 05-MAR-1993
; APPLICATION NUMBER: US 08/103,396
; FILING DATE: 06-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 018623-005030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX:
; INFORMATION FOR SEQ ID NO: 833:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-159-339A-833

Query Match 35.7%; Score 5; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AGFKG 10
Db 1 AGFKG 5

RESULT 9
PCT-US96-00206-2
; Sequence 2, Application PC/TUS9600206
; GENERAL INFORMATION:
; APPLICANT: Immulogic Pharmaceutical Corporation
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: TREATING RHEUMATOID ARTHRITIS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lappin & Kusmer
; STREET: 200 State Street
; CITY: Boston
; STATE: MA

; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00206
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: IMZ-014PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-466-6000
; TELEFAX: 617-466-6040
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus type II collagen
; PCT-US96-00206-2

Query Match 35.7%; Score 5; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AGFKG 10
Db 1 AGFKG 5

RESULT 10
PCT-US96-00206-3
; Sequence 3, Application PC/TUS9600206
; GENERAL INFORMATION:
; APPLICANT: Immulogic Pharmaceutical Corporation
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: TREATING RHEUMATOID ARTHRITIS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lappin & Kusmer
; STREET: 200 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00206
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: IMZ-014PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-466-6000
; TELEFAX: 617-466-6040
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids

```
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus type II collagen
PCT-US96-00206-3

Query Match 35.7%; Score 5; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AGFKG 10
Db 4 AGFKG 8

RESULT 11
PCT-US96-00206-4
; Sequence 4, Application PC/TUS9600206
; GENERAL INFORMATION:
; APPLICANT: Immunologic Pharmaceutical Corporation
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: TREATING RHEUMATOID ARTHRITIS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lappin & Kusmer
; STREET: 200 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/00206
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: INZ-014PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-466-6000
; TELEFAX: 617-466-6040
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus type II collagen
PCT-US96-00206-4

Query Match 35.7%; Score 5; DB 5; Length 16;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AGFKG 10
Db 4 AGFKG 8

RESULT 12
US-08-521-871A-5
; Sequence 5, Application US/08521871A

; Patent No. 5965787
; GENERAL INFORMATION:
; APPLICANT: Luthra, Harvinder S.
; APPLICANT: David, Chella S.
; APPLICANT: Zanelli, Eric
; TITLE OF INVENTION: HLA-DRB1 PEPTIDES WITH SPECIFIC BINDING
; TITLE OF INVENTION: AFFINITY FOR HLA-DQ MOLECULES: PREVENTION AND
; TITLE OF INVENTION: TREATMENT OF RHEUMATOID ARTHRITIS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C., P.A.
; STREET: 60 South Sixth Street, Suite 3300
; CITY: Minnesota
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/521.871A
; APPLICATION NUMBER: 08/31/95
; FILING DATE: 514
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellinger, Mark S.
; REGISTRATION NUMBER: 34,812
; REFERENCE/DOCKET NUMBER: 07039/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/335-5070
; TELEFAX: 612/288-9696
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-521-871A-5

Query Match 35.7%; Score 5; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AGFKG 10
Db 13 AGFKG 17

RESULT 13
US-07-951-565-1
; Sequence 1, Application US/07951565
; Patent No. 5399347
; GENERAL INFORMATION:
; APPLICANT: Trentham, David E.
; APPLICANT: Weiner, Howard L.
; TITLE OF INVENTION: Method of Treating Rheumatoid Arthritis with
; TITLE OF INVENTION: Type II Collagen
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/951,565
; FILING DATE: 19920925
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C.
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 1010/07300
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Gallus domesticus
; TISSUE TYPE: collagen
; US-07-951-565-1

Query Match 35.7%; Score 5; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKG 10
DB 17 AGFKG 21

RESULT 14
US-07-951-565-6
; Sequence 6, Application US/07951565
; Patent No. 5399347
; GENERAL INFORMATION:
; APPLICANT: Trentham, David E.
; APPLICANT: Weiner, Howard L.
; TITLE OF INVENTION: Method of Treating Rheumatoid Arthritis with
; TYPE OF INVENTION: Type II Collagen
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/951,565
; FILING DATE: 19920925
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C.
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 1010/07300
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Gallus domesticus
; TISSUE TYPE: collagen
; US-07-951-565-1

Query Match 35.7%; Score 5; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKG 10
DB 17 AGFKG 21

RESULT 15
US-08-246-242-9
; Sequence 9, Application US/08246242
; Patent No. 5675060
; GENERAL INFORMATION:
; APPLICANT: Benoist, C.
; APPLICANT: Mathis, D.
; APPLICANT: Kouskoff, V.
; TITLE OF INVENTION: Transgenic Arthritic Mice
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246,242
; FILING DATE: 19-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 1383.0080000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-246-242-9

Query Match 35.7%; Score 5; DB 1; Length 26;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKG 10
DB 17 AGFKG 21

Search completed: January 31, 2002, 13:15:12
Job time: 93 sec
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:07 ; Search time 140.03 seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-74

Perfect score: 14

Sequence: 1 AGWIDAGFKKITL 14

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_ll01.*

- 1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	19	AAW72850
2	6	42.9	82	21	AAW59072
3	6	42.9	255	22	AAW01029
4	6	42.9	260	22	AAW79890
5	6	42.9	266	22	AAW89903
6	6	42.9	332	21	AAW02857
7	6	42.9	509	22	AAW93339
8	6	42.9	673	22	AAW94067
9	6	42.9	750	22	AAW94129
10	5	35.7	7	21	AAW35994
11	5	35.7	13	17	AAW03105

12	5	35.7	13	17	AAW03106	Bovine type II col
13	5	35.7	13	21	AAW82065	Collagen II (CII)
14	5	35.7	13	21	AAV58994	Type II collagen p
15	5	35.7	13	22	AAW84074	Peptide associated
16	5	35.7	13	22	AAW84108	Amino acid sequenc
17	5	35.7	15	21	AAW82066	Collagen II (CII)
18	5	35.7	16	17	AAW03107	Bovine type II col
19	5	35.7	16	22	AAW96875	Rheumatoid arthrit
20	5	35.7	20	22	AAW96873	Rheumatoid arthrit
21	5	35.7	21	19	AAW53854	III 250-270 peptid
22	5	35.7	21	21	AAW23766	Entry vector pENTR
23	5	35.7	22	20	AAW26976	Human type II coll
24	5	35.7	25	20	AAW41477	Fragment of human
25	5	35.7	26	15	AAW51310	Type II collagen p
26	5	35.7	26	16	AAW98363	Arthritis toleroge
27	5	35.7	27	17	AAW03108	Bovine type II col
28	5	35.7	27	22	AAW84091	Immunomodulatory p
29	5	35.7	28	21	AAW39279	Gene 33 human secr
30	5	35.7	31	22	AAW84093	Immunomodulatory p
31	5	35.7	33	22	AAW84094	Immunomodulatory p
32	5	35.7	33	22	AAW84095	Immunomodulatory p
33	5	35.7	47	14	AAW33767	Wheat gammal purot
34	5	35.7	48	14	AAW33766	Sorghum alpha-amyl
35	5	35.7	51	21	AAW01170	Human secreted pro
36	5	35.7	53	16	AAW79479	Human type II coll
37	5	35.7	53	21	AAW12273	Peptide used in is
38	5	35.7	60	21	AAW18359	Arabidopsis thalia
39	5	35.7	66	22	AAW37414	Peptide #11451 enc
40	5	35.7	72	22	AAW14925	Novel bone marrow
41	5	35.7	75	21	AAW50217	Arabidopsis thalia
42	5	35.7	79	20	AAW36443	Fragment of human
43	5	35.7	83	20	AAW12895	Human 5' EST secre
44	5	35.7	85	21	AAW24263	Arabidopsis thalia
45	5	35.7	87	21	AAW16519	Arabidopsis thalia

ALIGNMENTS

RESULT	1
AAW72850	
ID	AAW72850 standard; Peptide: 14 AA.
AC	XX
AC	AAW72850;
XX	
DT	01-MAR-1999 (first entry)
XX	
DE	Methanococcus jannaschlii dCTP deaminase uridine-binding motif.
XX	
KW	Polymerase enhancing factor; PEF; dCTP deaminase; PCR;
KW	amplification; sequencing; replication.
XX	
OS	Methanococcus jannaschlii.
XX	
PN	W09842860-A1.
XX	
PD	01-OCT-1998.
XX	
PF	20-MAR-1998; 98WO-US05497.
XX	
PR	24-OCT-1997; 97US-0957709.
PR	21-MAR-1997; 97US-0822774.
XX	
PA	(STRA-) STRATAGENE.
PI	Hansen CJ, Hogrefe H;
XX	
DR	WPI; 1998-542284/46.
XX	
PT	Polymerase enhancing factor proteins, extracts and complexes -
PT	improve the polymerisation activity of nucleic acid polymerases, for
PT	use in amplification, sequencing and replication
XX	

Claim 71; Page 47; 16lpp; English.

This is the uridine-binding motif of the dCTP deaminase of *Methanococcus jannaschii*. Sequences are provided (see AAW72849-57) of the uridine-binding motifs of dUTPases and dCTP deaminases of *Pyrococcus furiosus* (see AAW72847), *Methanococcus jannaschii*, *Desulfurolobus ambivalens*, *Escherichia coli*, yeast, human and herpesvirus; a consensus (see AAW72848) is also provided. A claimed method of enhancing a nucleic acid polymerase reaction comprises performing the reaction in the presence of one or more of the following: a polymerase enhancing factor (PEF), a dUTPase, a protein that turns-over dUTP and a protein having one or more of the sequences provided in AAW72848-57. A claimed protein having PEF activity comprises one or more of sequences given in AAW72848-57. Kits are provided for replicating nucleic acids, for site-directed mutagenesis, for nucleic acid sequencing or for amplification (preferably PCR or RT-PCR).

Sequence 14 AA:

Query Match 100.0%; Score 14; DB 19; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGWIDAGFGKRTL 14
| | | | | | | | | | | | | |
Db 1 agwidagfgkrtl 14

RESULT 2

ID AAB59072 standard; Protein; 82 AA.

AC AAB59072;

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 780.

Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive; neutropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; immune disorder; Addison's disease; allergy; autoimmune haemolytic anaemia; autoimmune thyroiditis; diabetes mellitus; Crohn's disease; multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disorder; wound healing; neurological disease.

Homo sapiens.

WO200005173-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05881.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI: 2000-611515/58.

N-PSDB: AAF21975.

New human breast and ovarian cancer associated gene sequences and the polypeptides encoded by these genes, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases.

Claim 11; Page 1244; 1299pp; English.

Sequences AAF21614 - AAF22031 represent DNA sequences encoding human proteins AAB58711 - AAB59128. The DNA and protein sequences are associated with breast and ovarian cancer. Included in the invention are sequences AAF22032 - AAF22040 and AAB59129 which are used in the isolation and characterisation of the DNA and protein sequences of the invention. The breast and ovarian cancer associated DNA, protein, agonist or antagonist sequences exhibit cytostatic; immunosuppressive; neutropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic and cardiant activity. The polynucleotide and protein sequences are used in the diagnosis of cancer, particularly breast and ovarian cancer. The nucleic acid sequences, proteins, agonists and antagonists may also be used in the diagnosis, prevention and treatment of immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, multiple sclerosis, rheumatoid diabetes mellitus, Crohn's disease, cardiovascular disorders such as arthritis and ulcerative colitis; wound healing; neurological diseases such as myocardial ischaemias; wound healing; neurological diseases such as cerebral anoxia and epilepsy; and infectious diseases.

Sequence 82 AA:

Query Match 42.9%; Score 6; DB 21; Length 82;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 FKGGKIT 13
| | | | | | | | | | | | | |
Db 15 fkgkkit 20

RESULT 3

ID AAM01029

AC AAM01029 standard; Protein; 255 AA.

AC AAM01029;

DT 02-OCT-2001 (first entry)

DE CFE 32 protein sequence.

Antibacterial; vaccine; gene therapy; bacterial cell wall viability; CFE; CEG; Conserved Essential Gene; bacterial infection; antisense therapy; antibiotic resistance.

Streptococcus pneumoniae.

WO200149721-A2.

12-JUL-2001.

29-DEC-2000; 2000WO-US35604.

30-DEC-1999; 99US-0174089.

(BRIM) BRISTOL-MYERS SQUIBB CO.

Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Brucoleri RE; Thanassi JA;

WPI: 2001-496721/54.

N-PSDB: AAB90728.

Nucleic acids encoding conserved essential genes involved in bacterial replication which are potential targets for the treatment of antibiotic resistant bacterial infections.

Claim 27; Pages 270-271; 380pp; English.

The present invention relates to nucleic acids (AAB90701-AAH90918) encoding polypeptides (AAM01002-AAM01114), which are essential for the viability of a bacterial cell wall. The acronym CFE stands for "CEG For

CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic
 CC acids are useful for detecting the presence of proteins essential for the
 CC viability of a bacterial cell wall in samples such as cells, tissues,
 CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,
 CC and for detecting corresponding target nucleic acid molecules with
 CC complementary sequences. The nucleic acids are also useful for
 CC determining whether a genomic nucleotide sequence of interest is
 CC essential for viability of a bacterial cell or whether it resides within
 CC an operon, by integrating an exogenous nucleotide sequence comprising a
 CC portion of an open reading frame of the genomic sequence of interest
 CC (comprising 200-500 base pairs) into the genomic sequence of interest
 CC which confers a selectable agent to the cell, and determining cell
 CC viability with a selection agent such as chloramphenicol. The nucleic
 CC acids and proteins are also useful as vaccines and for treating bacterial
 CC infections with gene therapy and antisense therapy. The nucleic acids
 CC also enable identification of targets suitable for the treatment of
 CC antibiotic resistant bacterial infections.

XX Sequence 255 AA;

Query Match 42.9%; Score 6; DB 22; Length 255;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 AGFKGK 11

Db 8 agfkgk 13

RESULT 4

AAAB79890
 ID AAB79890 standard; Protein; 260 AA.

XX

AC AAB79890;

XX

DT 30-APR-2001 (first entry)

XX

DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:514.

XX

KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;

KW fine chemical production; microorganism; organic acid; nucleoside;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.

XX

OS Corynebacterium glutamicum.

XX

PN WO200100843-A2.

XX

PD 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-IB00923.

XX

PR 25-JUN-1999; 99US-0141031.

PR 01-JUL-1999; 99DE-1030476.

PR 02-JUL-1999; 99US-0142101.

PR 08-JUL-1999; 99DE-1031415.

PR 08-JUL-1999; 99DE-1031418.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031435.

PR 08-JUL-1999; 99DE-1031443.

PR 08-JUL-1999; 99DE-1031453.

PR 08-JUL-1999; 99DE-1031457.

PR 08-JUL-1999; 99DE-1031465.

PR 08-JUL-1999; 99DE-1031478.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031541.

PR 08-JUL-1999; 99DE-1031573.

PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 08-JUL-1999; 99DE-1032125.
 PR 08-JUL-1999; 99DE-1032126.
 PR 08-JUL-1999; 99DE-1032130.
 PR 08-JUL-1999; 99DE-1032186.
 PR 08-JUL-1999; 99DE-1032206.
 PR 08-JUL-1999; 99DE-1032227.
 PR 08-JUL-1999; 99DE-1032228.
 PR 08-JUL-1999; 99DE-1032229.
 PR 08-JUL-1999; 99DE-1032230.
 PR 08-JUL-1999; 99DE-1032232.
 PR 08-JUL-1999; 99DE-1032926.
 PR 08-JUL-1999; 99DE-1032928.
 PR 08-JUL-1999; 99DE-1033004.
 PR 08-JUL-1999; 99DE-1033005.
 PR 08-JUL-1999; 99DE-1033006.
 PR 08-JUL-1999; 99US-0148613.
 PR 08-JUL-1999; 99DE-1040764.
 PR 08-JUL-1999; 99DE-1040765.
 PR 08-JUL-1999; 99DE-1040766.
 PR 08-JUL-1999; 99DE-1040832.
 PR 08-JUL-1999; 99DE-1041378.
 PR 08-JUL-1999; 99DE-1041379.
 PR 08-JUL-1999; 99DE-1041380.
 PR 08-JUL-1999; 99DE-1041394.
 PR 08-JUL-1999; 99DE-1041396.
 PR 08-JUL-1999; 99DE-1042076.
 PR 08-JUL-1999; 99DE-1042077.
 PR 08-JUL-1999; 99DE-1042079.
 PR 08-JUL-1999; 99DE-1042086.
 PR 08-JUL-1999; 99DE-1042087.
 PR 08-JUL-1999; 99DE-1042088.
 PR 08-JUL-1999; 99DE-1042095.
 PR 08-JUL-1999; 99DE-1042124.
 PR 08-JUL-1999; 99DE-1042129.
 PR 08-JUL-1999; 2000US-0187970.

(BADI) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

WPI: 2001-137957/14.

N-PSDB: AAF72009.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 42.9%; Score 6; DB 22; Length 260;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGWIDA 6

Db 20 agwida 25

Sequence 260 AA;

Nucleic acids from Corynebacterium glutamicum encoding metabolic
 pathway proteins, useful for producing fine chemicals in
 microorganisms, including organic acids, nonproteinogenic amino acids,
 and purine and pyrimidine bases -

Claim 20; Page 896-897; 1737pp; English.

AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 MP nucleic acids are useful for the production of fine chemicals
 in microorganisms, including organic acids, nonproteinogenic amino
 acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 compounds, vitamins, cofactors, polyketides and enzymes.

RESULT 5
 AAC89903
 ID AAC89903 standard; Protein: 266 AA.
 AC AAC89903;
 DT 26-SEP-2001 (first entry)
 DE C glutamicum protein fragment SEQ ID NO: 3657.
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 OS Corynebacterium glutamicum.
 PN EPI108790-A2.
 PD 20-JUN-2001.
 PF 18-DEC-2000; 2000EP-0127688.
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI: 2001-376931/40.
 DR N-PSDB: AAH65122.
 DR Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 XX Claim 17: SEQ ID NO: 3657; 246pp + Sequence Listing; English.
 PS The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium, and identifying a homologue of a gene derived
 CC from Corynebacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 266 AA;

Query Match 42.9%; Score 6; DB 22; Length 266;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGWTDA 6
 Db 26 agwida 31

RESULT 6
 AAB02857
 ID AAB02857 standard; Protein: 332 AA.
 AC AAB02857;
 XX

DT 22-AUG-2000 (first entry)
 XX Human G protein coupled receptor hmc4 (A244K) protein SEQ ID NO:136.
 DE Human; G protein coupled receptor; GPCR; transmembrane receptor;
 KW identification; agonist; screening; therapeutic; pharmaceutical;
 KW mutant.
 OS Homo sapiens.
 OS Synthetic.
 PN WO2000022131-A2.
 XX 20-APR-2000.
 PD 13-OCT-1999; 99WO-US24065.
 PF 13-OCT-1998; 98US-0170496.
 PR 12-NOV-1998; 98US-0108029.
 PR 20-NOV-1998; 98US-0109213.
 PR 27-NOV-1998; 98US-0110060.
 PR 16-FEB-1999; 98US-0120416.
 PR 26-FEB-1999; 98US-0121852.
 PR 12-MAR-1999; 98US-0123944.
 PR 12-MAR-1999; 98US-0123945.
 PR 12-MAR-1999; 98US-0123946.
 PR 12-MAR-1999; 98US-0123948.
 PR 12-MAR-1999; 98US-0123951.
 PR 28-MAY-1999; 98US-0136436.
 PR 28-MAY-1999; 98US-0136437.
 PR 28-MAY-1999; 98US-0137127.
 PR 28-MAY-1999; 98US-0137131.
 PR 28-MAY-1999; 98US-0137567.
 PR 30-JUN-1999; 98US-0141448.
 PR 27-AUG-1999; 98US-0151114.
 PR 03-SEP-1999; 98US-0152524.
 PR 29-SEP-1999; 98US-0156633.
 PR 29-SEP-1999; 98US-0156555.
 PR 29-SEP-1999; 98US-0156634.
 XX (AREN-) ARENA PHARM INC.
 XX Behan DP, Lehmann-Bruinsma K, Chalmers DT, Chen R, Dang HT;
 PI Gore M, Liaw CW, Lin I, Lowitz K, White C;
 XX WPI: 2000-317986/27.
 DR N-PSDB: AAA46119.
 DR Non-endogenous, human G protein-coupled receptors for screening
 XX receptor, inverse or partial agonists useful as therapeutic agents
 PT Example 2; Page 176-177; 187pp; English.
 XX The present invention describes transmembrane receptors, preferably
 CC human G protein coupled receptors (GPCR), for which the endogenous
 CC ligand is unknown (orphan GPCR receptors). More specifically the present
 CC invention relates to non-endogenous, constitutively activated versions
 CC of a human GPCR. These non-endogenous human GPCRs can be useful for
 CC the direct identification of candidate compounds as receptor agonists,
 CC inverse agonists or partial agonists for use as pharmaceutical agents.
 CC AAA46017 to AAA46126 and AAB02825 to AAB02859 represent sequences used in
 CC the exemplification of the present invention.
 XX
 SQ Sequence 332 AA;

Query Match 42.9%; Score 6; DB 21; Length 332;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 KGRITL 14

Best Local Similarity 100.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11
| | | | |
Db 533 agfkkgk 538

RESULT 9
AAB94129
ID AAB94129 standard; Protein; 750 AA.
XX AC AAB94129;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:14385.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EPI074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118752.
XX PR 02-MAY-2000; 2000JP-0183769.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 14385; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides, and the combination
XX of the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences. AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

XX Sequence 750 AA;

Query Match 35.7%; Score 5; DB 21; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DAGFK 9
| | | | |
Db 3 dagfk 7

RESULT 11
AAW03105

Query Match 42.9%; Score 6; DB 22; Length 750;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKGK 11
| | | | |
Db 606 agfkkgk 611

RESULT 10
AAB35994
ID AAB35994 standard; Protein; 7 AA.
XX AC AAB35994;
XX DT 01-MAR-2001 (first entry)
XX DE Sorbitol dehydrogenase subunit 2 internal peptide.
XX KW Sorbitol dehydrogenase; SDH; Gluconobacter oxydans; D-sorbitol;
XX L-sorbose production; 2-keto-L-gulononic acid.
XX OS Gluconobacter oxydans.
XX PN WO2000065066-A1.
XX PD 02-NOV-2000.
XX PF 23-APR-1999; 99WO-IB00736.
XX PR 23-APR-1999; 99WO-IB00736.
XX PA (CHOI/) CHOI E.
XX PA (RHEE/) RHEE S.
XX PA (LEEE/) LEE E.
XX PI Choi E, Rhee S, Lee E;
XX WPI: 2000-687351/67.
XX
XX Novel membrane-bound sorbitol dehydrogenase nucleic acid molecule
XX isolated from Gluconobacter suboxydans useful for the fermentative
XX production of 2-keto-L-gulononic acid and L-sorbose from D-sorbitol -
XX
XX Example 2; Page 37; 96pp; English.
XX
XX This invention relates to an isolated membrane-bound sorbitol
XX dehydrogenase (SDH) from Gluconobacter oxydans. The invention includes
XX polynucleotides (AAC83153 - AAC83154) encoding the 3 subunits of SDH
XX (AAB35987 - AAB35989). Also included in the invention are two
XX polynucleotide sequences AAC83156 and AAC83157 which contain the subunit
XX coding regions. The sorbitol dehydrogenase polynucleotide sequences are
XX useful for producing L-sorbose from D-sorbitol and for increasing the
XX production of 2-keto-L-gulononic acid by transforming a host cell,
XX especially Gluconobacter with the DNA and selecting the transformed host
XX cell. The present sequence represents the SDH subunit 2 internal
XX peptide.

XX Sequence 7 AA;

ID AAW03105 standard; peptide; 13 AA.

XX AC AAW03105;

XX DT 03-MAR-1997 (first entry)

XX DE Bovine type II collagen peptide (276-288).

XX KW Collagen; type II; Bovine; Human; rheumatoid arthritis; epitope;
human major histocompatibility complex; genetically linked.

XX OS Bos taurus.

XX PN W09620950-A2.

XX PD 11-JUL-1996.

XX PF 04-JAN-1996; 96WO-US00206.

XX PR 06-JAN-1995; 95US-0369792.

XX PA (IMMU-) IMMULOGIC PHARM CORP.

XX PI Fugger LH, Rothbard J, Sonderstrup-Mcdevitt G;

XX DR WPI; 1996-333937/33.

XX PT New peptide fragments from human type II collagen - bind to specific
PT major histocompatibility complex proteins and are useful, opt. with
PT known collagen fragments, to treat rheumatoid arthritis

XX PS Claim 1; Page 29; 46pp; English.

XX CC The present invention provides peptides, therapeutic compositions, and
CC methods for treatment of rheumatoid arthritis in mammals, specifically
CC in humans. The peptides of the invention comprise fragments of type II
CC collagen which bind specifically with human major histocompatibility
CC complex proteins known to be genetically linked to susceptibility to
CC rheumatoid arthritis. The therapeutic compositions of the invention
CC comprise the peptides, alone or in combination with other collagen
CC peptides. AAW03105-107 are claimed peptides which can be used to treat
CC rheumatoid arthritis by down-regulating the autoimmune response, esp.
CC rendering T cells non-responsive to the rheumatoid arthritis-related
CC autoantigen.

XX SQ Sequence 13 AA;

Query Match 35.7%; Score 5; DB 17; Length 13;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKG 10

Db 1 agfg 5

RESULT 12

AAW03106

ID AAW03106 standard; peptide; 13 AA.

XX AC AAW03106;

XX DT 03-MAR-1997 (first entry)

XX DE Bovine type II collagen peptide (273-285).

XX KW Collagen; type II; Bovine; Human; rheumatoid arthritis; epitope;
human major histocompatibility complex; genetically linked.

XX OS Bos taurus.

XX PN W09620950-A2.

XX PD 11-JUL-1996.

XX PF 04-JAN-1996; 96WO-US00206.

XX PR 06-JAN-1995; 95US-0369792.

XX PA (IMMU-) IMMULOGIC PHARM CORP.

XX PI Fugger LH, Rothbard J, Sonderstrup-Mcdevitt G;

XX DR WPI; 1996-333937/33.

XX PT New peptide fragments from human type II collagen - bind to specific
PT major histocompatibility complex proteins and are useful, opt. with
PT known collagen fragments, to treat rheumatoid arthritis

XX PS Claim 1; Page 30; 46pp; English.

XX CC The present invention provides peptides, therapeutic compositions, and
CC methods for treatment of rheumatoid arthritis in mammals, specifically
CC in humans. The peptides of the invention comprise fragments of type II
CC collagen which bind specifically with human major histocompatibility
CC complex proteins known to be genetically linked to susceptibility to
CC rheumatoid arthritis. The therapeutic compositions of the invention
CC comprise the peptides, alone or in combination with other collagen
CC peptides. AAW03105-107 are claimed peptides which can be used to treat
CC rheumatoid arthritis by down-regulating the autoimmune response, esp.
CC rendering T cells non-responsive to the rheumatoid arthritis-related
CC autoantigen.

XX SQ Sequence 13 AA;

Query Match 35.7%; Score 5; DB 17; Length 13;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFKG 10

Db 4 agfg 8

RESULT 13

AAW02065

ID AAW02065 standard; peptide; 13 AA.

XX AC AAW02065;

XX DT 01-JUN-2000 (first entry)

XX DE Collagen II (CII) peptide 261-273 SEQ ID NO:2.

XX KW MHC class II; major histocompatibility complex; autoimmune disease;
inflammatory disease; binding; rheumatoid arthritis; antiinflammatory;
XX KW antiarthritic; multiple sclerosis.

XX OS Synthetic.

XX PN W0200005249-A2.

XX PD 03-FEB-2000.

XX PF 22-JUL-1999; 99WO-US16617.

XX PR 23-JUL-1998; 98US-0093859.

XX PR 09-MAR-1999; 99US-0123675.

XX PA (HARD) HARVARD COLLEGE.

XX PI Strominger JL, Fridkis-Harell M;

XX DR WPI; 2000-205374/18.

XX New synthetic peptide, useful for treating autoimmune disease, e.g.
PT rheumatoid arthritis -
XX
XX
PS Example 1; Page 19; 57pp; English.
XX
XX The present invention describes synthetic peptides having an amino acid
CC sequence comprising at least 3 residues selected from the group of amino
CC acids, consisting of aromatic acids, negatively charged amino acids,
CC positively charged amino acids, and aliphatic amino acids, the synthetic
CC peptides being at least 7 amino acid residues in length and capable of
CC binding to a major histocompatibility complex (MHC) class II protein
CC associated with an autoimmune disease. The synthetic peptides have
CC anti-inflammatory and anti-arthritis activities. They are used to treat
CC rheumatoid arthritis and demyelinating autoimmune diseases, especially
CC for particular MHC Class II alleles. Purified, short and synthetic
CC peptides should have fewer side effects than mixtures of random
CC peptides; may include many repeats of the active sequence and/or contain
CC amino acid analogues that improve stability (or other desired features).
CC AAY82021 to AAY82044 represent specifically claimed peptide sequences
CC which can be used as part of the synthetic peptides of the present
CC invention; AAY82045 to AAY82063 represent specifically claimed examples
CC of the synthetic peptides from the present invention; and AAY82064 to
CC AAY82080 represent other peptides used in the exemplification of the
CC present invention.
XX
XX Sequence 13 AA;
XX
XX Query Match 35.7%; Score 5; DB 21; Length 13;
XX Best Local Similarity 100.0%; Pred. No. 38;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 6 AGFKG 10
XX | | | |
XX Db 1 agfkG 5
XX
XX RESULT 14
XX AAY58994
XX ID AAY58994 standard; Peptide; 13 AA.
XX AC AAY58994;
XX DT 23-MAY-2000 (first entry)
XX DE Type II collagen peptide 261-273.
XX
XX Collagen; antigen; autoimmune disease; multiple sclerosis;
XX autoimmune haemolytic anaemia; autoimmune oophoritis;
XX autoimmune thyroiditis; colitis; autoimmune uveoretinitis;
XX chronic immune thrombocytopenic purpura;
XX contact sensitivity disease; diabetes mellitus; Graves disease;
XX Guillain-Barre's syndrome; Hashimoto's disease;
XX idiopathic myxedema; myasthenia gravis; psoriasis;
XX pemphigus vulgaris; rheumatoid arthritis;
XX systemic lupus erythematosus; immunosuppressant; neuroprotective;
XX anti-anaemic; antithyroid; antidiabetic; thyromimetic;
XX antipsoriatic; antirheumatic; antiarthritis; dermatological;
XX antiinflammatory; therapy; major histocompatibility complex;
XX MHC class II; human lymphocyte antigen; HLA-DR.
XX
XX Unidentified.
XX OS
XX WO200005250-A1.
XX PN
XX PD 03-FEB-2000.
XX
XX 23-JUL-1999; 99WO-US16747.
XX PF
XX 23-JUL-1998; 98US-0093859.
XX PR
XX 25-SEP-1998; 98US-010825.
XX PR-

PR 02-OCT-1998; 98US-0102960.
PR 12-NOV-1998; 98US-0108184.
PR 09-MAR-1999; 99US-0123675.
XX
XX (YEDA) YEDA RES & DEV CO LTD.
XX (HARD) HARVARD COLLEGE.
XX
XX Aharoni R, Teitelbaum D, Arnon R, Sela M, Fridkis-Hareli M;
XX Strominger JI;
XX WPI; 2000-182641/16.
XX
XX New terpolymers, copeptides and copolymer I which contain three amino
XX acids randomly joined in a linear array where one is aromatic, one is
XX aliphatic and the other is charged, used to treat autoimmune diseases
XX -
XX
XX Example 11; Page 67; 147pp; English.
XX
XX The present sequence represents type II collagen peptide 261-273.
XX The peptide was used in the design of copeptides (see AAY58956-88)
XX that show a high affinity for MHC class II proteins associated with
XX an autoimmune disease, especially HLA-DRI, HLA-DR2 or HLA-DR4, bind
XX to antigen presenting cells, and inhibit T cell responses. The
XX copeptides are used to treat multiple sclerosis, autoimmune
XX haemolytic anaemia, autoimmune oophoritis, autoimmune thyroiditis,
XX autoimmune uveoretinitis, chronic immune thrombocytopenic purpura,
XX colitis, contact sensitivity disease, diabetes mellitus, Graves
XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
XX myxedema, myasthenia gravis, psoriasis, pemphigus vulgaris,
XX rheumatoid arthritis and systemic lupus erythematosus (all
XX claimed).
XX
XX Sequence 13 AA;
XX
XX Query Match 35.7%; Score 5; DB 21; Length 13;
XX Best Local Similarity 100.0%; Pred No 38;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 6 AGFKG 10
XX | | | |
XX Db 1 agfkG 5
XX
XX RESULT 15
XX AAB84074
XX ID AAB84074 standard; peptide; 13 AA.
XX AC AAB84074;
XX DT 06-AUG-2001 (first entry)
XX DE Peptide associated with rheumatoid arthritis type II.
XX
XX Immunomodulatory peptide; immune response; autoimmune disease; allergy;
XX asthma; host-versus-graft rejection; T cell; anergy; apoptosis.
XX
XX Unidentified.
XX OS
XX WO200136448-A2.
XX PN
XX 25-MAY-2001.
XX PD
XX 27-OCT-2000; 2000WO-US41646.
XX PF
XX 27-OCT-1999; 99US-0161734.
XX PR
XX (CELS-) CEL-SCI CORP.
XX PA
XX Zimmerman DH;
XX PI
XX WPI; 2001-374498/39.
XX DR

XX Novel immunomodulatory peptide construct useful for modulating an
PT inappropriate immune response in an individual at risk for autoimmune
PT disease, allergic reactions, asthma or host-graft or graft-host disease
PT
XX
XX
PS Claim 2; Page 36; 55pp; English.
XX
XX The specification describes an immunomodulatory peptide construct.
CC The immunomodulatory peptide comprises a first peptide associated with
CC autoimmune disease, allergy or asthma, or host-versus-graft rejection
CC and which will bind to an antigen receptor on a set or subset of T cells,
CC linked to a second immune modulating peptide which will cause a directed
CC immune response by the set or subset of T cells to which the first
CC peptide is attached. Alternatively, the second peptide will bind to a
CC T cell receptor site on the surface of the T cell which will cause the
CC set or subset of T cells to initiate, but not complete, an immune
CC response to cause the set or subset of T cells to undergo anergy and
CC apoptosis. The immunomodulatory peptides are useful for eliminating a set
CC or subset of T cells involved in autoimmune response. They are useful for
CC the treatment of autoimmune disease, allergic reactions, asthma or
CC host-graft or graft-host rejections. The immunomodulatory peptides are
CC also useful for interrupting an autoimmune disease associated pathway
CC necessary to complete T cell activation. The present peptide is used to
CC construct immunomodulatory peptides of the invention, and is
XX representative of the first peptide.
XX
SQ Sequence 13 AA;

Query Match 35.7%; Score 5; DB 22; Length 13;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 AGFKG 10
| | | | |
Db 2 agfk 6

Search completed: January 31, 2002, 13:18:08
Job time: 169 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:42 ; Search time 130.99 Seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-75
Perfect score: 14
Sequence: 1 SAVHDPGYEGREY 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL17:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	161	Q58502	Q58502 methanococ
2	7	50.0	1197	Q69101	Q69101 herpes simp
3	6	42.9	259	Q86610	Q86610 streptomyc
4	6	42.9	473	Q9AE03	Q9AE03 ancycolatops
5	6	42.9	487	Q9BTG9	Q9BTG9 homo sapien
6	6	42.9	609	Q9FAK2	Q9FAK2 streptomyc
7	6	42.9	634	Q9H2F4	Q9H2F4 pseudomonas
8	6	42.9	653	Q9A2E5	Q9A2E5 caulobacter
9	6	42.9	1498	Q9P884	Q9P884 emericella
10	5	35.7	40	Q18923	Q18923 sus scrofa
11	5	35.7	40	Q9FWM0	Q9FWM0 gallus gall
12	5	35.7	42	Q18958	Q18958 bos taurus
13	5	35.7	69	Q04128	Q04128 zea mays (m
14	5	35.7	101	Q9QVS3	Q9QVS3 rattus sp.
15	5	35.7	106	Q9JIW2	Q9JIW2 rattus norv
16	5	35.7	118	Q9BGP0	Q9BGP0 macaca fasc
17	5	35.7	134	Q9SSB5	Q9SSB5 desulfovibr
18	5	35.7	136	Q30350	Q30350 salmonella
19	5	35.7	139	Q51623	Q51623 escherichia

20	5	35.7	140	2	P72692	P72692 synechocyst
21	5	35.7	140	2	Q9S2G8	Q9S2G8 streptomyc
22	5	35.7	142	10	O82701	O82701 medicago tr
23	5	35.7	147	7	Q31473	Q31473 nerodia sip
24	5	35.7	149	2	O68068	O68068 rhodobacter
25	5	35.7	149	4	O9H6Y4	O9H6Y4 homo sapien
26	5	35.7	149	10	Q01737	Q01737 phytophthor
27	5	35.7	154	10	Q39814	Q39814 glycine max
28	5	35.7	155	10	Q12536	Q12536 achlya ambi
29	5	35.7	161	2	P74069	P74069 synechocyst
30	5	35.7	161	12	O64778	O64778 avian adeno
31	5	35.7	164	2	O53952	O53952 streptomyc
32	5	35.7	165	2	O9K3R4	O9K3R4 streptomyc
33	5	35.7	167	7	O46731	O46731 bos taurus
34	5	35.7	167	7	O46719	O46719 bos taurus
35	5	35.7	167	7	O46754	O46754 bos taurus
36	5	35.7	167	7	O46755	O46755 bos taurus
37	5	35.7	167	7	O46756	O46756 bos taurus
38	5	35.7	167	7	O46762	O46762 bos taurus
39	5	35.7	168	1	O29157	O29157 archaeoglob
40	5	35.7	178	3	P78737	P78737 pyrenophora
41	5	35.7	178	10	Q9FTT6	Q9FTT6 oryza sativ
42	5	35.7	181	2	Q9ZAD5	Q9ZAD5 porphyromon
43	5	35.7	185	2	Q9ZET9	Q9ZET9 mycobacteri
44	5	35.7	190	10	O65637	O65637 arabidopsis
45	5	35.7	191	2	Q9RVE7	Q9RVE7 deinococcus

ALIGNMENTS

RESULT	1				
Q58502					
ID	Q58502	PRELIMINARY;	PRT:	161	AA.
AC	Q58502;				
DT	01-JAN-1998 (TRENBLrel. 05, Created)				
DT	01-JAN-1998 (TRENBLrel. 05, Last sequence update)				
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)				
DE	HYPOTHETICAL PROTEIN MJ1102.				
GN	MJ1102.				
OS	Methanococcus jannaschii.				
OC	Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;				
OC	Methanococcus.				
OX	NCBI_TaxID=2190;				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RP	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;				
RX	MEDLINE=96337999; PubMed=8688087;				
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,				
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Adams M.D., Reich C.I.,				
RA	Kerlavage A.R., Dougherty B.A., Tomb J.F., Merrick J.M., Glodek A.,				
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,				
RA	Scott J.L., Geohagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,				
RA	Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,				
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,				
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;				
RT	*Complete genome sequence of the methanogenic archaeon, Methanococcus				
RL	jannaschii.;				
CC	Science 273:1058-1073(1996).				
DR	-I- SIMILARITY: TO BACTERIAL DEOXYCYTIDINE TRIPHOSPHATE DEAMINASE.				
DR	EMBL; U67553; AAB9105.1; -				
DR	TIGR; MJ1102.				
DR	InterPro; IPR003232; dCTP_deaminase.				
DR	InterPro; IPR001428; dUTPase.				
DR	Pfam; PF00692; dUTPase; 1.				
DR	ProDom; PD004900; dCTP_deaminase; 1.				
KW	Hypothetical protein; Complete proteome.				
SK	SEQUENCE 161 AA; 18640 MW; 886357CB4A533CF4 CRC64;				

Query Match 100.0%; Score 14; DB 1; Length 161;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SAVHDPGYEGRP 14
 Db 110 SAVHDPGYEGRP 123

RESULT 2

ID Q69101 PRELIMINARY; PRT; 1197 AA.
 AC Q69101;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE DNA BINDING PROTEIN ICP8.
 DE Herpes simplex virus (type 2).
 OS Herpes simplex virus (type 2).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxID=10310;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KN;
 RX MEDLINE=93228441; PubMed=8385914;
 RA Toh Y., Tanaka S., Liu Y., Mori R.;
 RT "Nucleotide sequence of the major DNA-binding protein gene of herpes
 simplex virus type 2 and a comparison with the type 1.";
 RL Arch. Virol. 128:183-196(1993).
 DR EMBL: D10658; BAA01507.1;
 DR InterPro: IPR000635; Viral_DNA_bind.
 DR Pfam: PF00747; Viral_DNA_dp; 1.
 DR Q69101; 1197 AA; 128470 MW; AA3ADA75B8865BFE CRC64;
 SQ SEQUENCE 1197 AA; Score 7; DB 12; Length 1197;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 100.0%; Pred. No. 8;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 PGYEGRP 12
 Db 429 PGYEGRP 435

RESULT 3

ID O86610 PRELIMINARY; PRT; 259 AA.
 AC O86610;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PUTATIVE EXONUCLEASE.
 GN SC3A7.09.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kleser H.M., Denapite D., Eichner A., Cullum J.;
 RL MEDLINE=9700351; PubMed=8843436;
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).

DR EMBL: AL031155; CAA20072.1; -.
 DR HSP: P09030; IAKO.
 DR InterPro: IPR000097; AP_endonclse_family_1.
 DR Pfam: PF01260; AP_endonuclease1; 1.
 KW Exonuclease.
 SQ SEQUENCE 259 AA; 28431 MW; 11C0DA5CD6FC6B58 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 259;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 DPGYEG 10
 Db 83 DPGYEG 88

RESULT 4

ID Q9AE03 PRELIMINARY; PRT; 473 AA.
 AC Q9AE03;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE RIF18.
 GN RIF18.
 OS Amycolatopsis mediterranei.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
 OX NCBI_TaxID=33910;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=98174059; PubMed=9512878;
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Yu T.W.,
 RA Taylor M., Hoffmann D., Kim C.G., Zhang X., Hutchinson C.R.,
 RA Floss H.G.;
 RT "Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
 the molecular analysis of the rif biosynthetic gene cluster of
 Amycolatopsis mediterranei S699.";
 RL Chem. Biol. 5:69-79(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=98165773; PubMed=9497318;
 RA Kim C.G., Yu T.W., Fryhle C.B., Handa S., Floss H.G.;
 RT "3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
 formation of the precursor of mC7N units in rifamycin and related
 antibiotics.";
 RL J. Biol. Chem. 273:6030-6040(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RX MEDLINE=21201076; PubMed=11278540;
 RA Yu T.W., Mueller R., Mueller M., Zhang X., Draeger G., Kim C.-G.,
 RA Leistner E., Floss H.G.;
 RT "Mutational analysis and reconstituted expression of the biosynthetic
 genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the
 starter unit of rifamycin biosynthesis in Amycolatopsis mediterranei
 S699.";
 RL J. Biol. Chem. 276:12546-12555(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S699;
 RA August P.R., Tang L., Yoon Y.J., Ning S., Mueller R., Hutchinson C.R.,
 RA Floss H.G.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF040570; AAG52988.1;
 SQ SEQUENCE 473 AA; 53211 MW; 6DE809458209BF2C CRC64;

Query Match 42.9%; Score 6; DB 2; Length 473;
 Best Local Similarity 100.0%; Pred. No. 46;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AVHDPG 7

Db 261 AVHDPG 266

|||||

RESULT 5

O9BTG9

ID O9BTG9

AC O9BTG9

DT 01-JUN-2001

DT 01-JUN-2001

DE UNKNOWN

OS Homo sapiens

OC Eukaryota

OC Mammalia

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LYMPHOMA;

RA Strausberg R.

RL Submitted (MAR-2001)

DR EMBL; BC004108; AA04108.1;

FT NON_TER

SQ SEQUENCE 487 AA; 51285 MW; EF643482D26BF14D CRC64;

Query Match

Best Local Similarity 42.9%; Score 6; DB 4; Length 487;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 PGYEGR 11

Db 246 PGYEGR 251

|||||

RESULT 6

O9FAR2

ID O9FAR2

AC O9FAR2

DT 01-MAR-2001

DT 01-MAR-2001

DE ASPARAGINE SYNTHASE

GN TCSG.

OS Streptomyces aureofaciens.

OC Bacteria; Firmicutes; Actinobacteri

OC Actinomycetales; Streptomycetaceae;

OX NCBI_TaxID=1894;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NRRL3203;

RA Nakano T., Ikeda M., Mizukami T.,

RT "Analysis of the chlorotetracycline

RL Streptomyces aureofaciens NRRL3203.

DR EMBL; AB039379; BAB12569.1;

DR InterPro; IPR001962; Asn_synthase.

DR Pfam; PF00733; Asn_synthase_2.

DR Pfam; PF00733; Asn_synthase_3.

DR PROSITE; PS00443; GATASE_TYPE_II;

SQ SEQUENCE 609 AA; 67138 MW; 9BFB791CD9586D65 CRC64;

Query Match

Best Local Similarity 42.9%; Score 6; DB 2; Length 609;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 DPGYEG 10

Db 5 DPGYEG 10

|||||

Db 539 DPGYEG 544

RESULT 7

O9HZF4

ID O9HZF4

AC O9HZF4

DT 01-MAR-2001

DT 01-MAR-2001

DE HYPOTHETICAL PROTEIN

GN PA3054.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma

OC Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PA01;

RX MEDLINE=2043737; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle

RA Garber R.L., Coltry L., Tolentino E.,

RA Brody L.L., Coulter S.H., Folger K.R.,

RA Smith K.A., Spencer D.H., Wong G.K.-S.,

RA Reizer J., Sailer M.H., Hancock R.E.W.,

RT "Complete genome sequence of

RT opportunistic pathogen";

RL Nature 406:959-964(2000).

DR EMBL; AE004730; AAG06442.1;

KW Hypothetical protein; Complete

SQ SEQUENCE 634 AA; 71197 MW; FE2043295E3770F9 CRC64;

Query Match

Best Local Similarity 42.9%; Score 6; DB 2; Length 634;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 PGYEGR 11

Db 504 PGYEGR 509

|||||

RESULT 8

O9A2E5

ID O9A2E5

AC O9A2E5

DT 01-JUN-2001

DT 01-JUN-2001

DE TRANSKETOLASE I.

GN CC3620.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha

OC Caulobacter.

OX NCBI_TaxID=69394;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21173698; PubMed=11259647;

RA Nierman W.C., Feldblyum T.V., Laub

RA Eisen J., Heidelberg J.F., Alley M.R.K.,

RA Potocka I., Nelson W.C., Newton A.,

RA DeBoy R.T., Dodson R.J., Durkin A.S.,

RA Kolony J.F., Smit J., Craven M.B.,

RA Utterback T., Tran K., Wolf A.,

RA Salzberg S.L., Venter J.C., Shapiro

RT "Complete genome sequence of

RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-

DR EMBL; AE006020; AAK25582.1;

DR Tigr; CC3620;

KW Complete proteome.

SQ SEQUENCE 653 AA; 69370 MW; 2E0CF8479A561C61 CRC64;

```

Query Match      42.9%; Score 6; DB 2; Length 653;
Best Local Similarity 100.0%; Pred. No. 61; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 6 PYEGR 11
DB 392 PYEGR 397

RESULT 9
Q9P884 Q9P884 PRELIMINARY; PRT; 1498 AA.
AC Q9P884;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ABC TRANSPORTER PROTEIN.
GN ATRE.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-WG096;
RA Andrade A.C., Van Nistelrooy J.G.M., Haas H., De Waard M.A.;
RT "ABC transporters and resistance to azole fungicides in the ima
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DE EMBL: AJ276241; CAB76823.1; -.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003439; ABC transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR000847; HTH_LysR.
DR InterPro: IPR001063; Ribosomal_L22.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN 1.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN 1.
DR PROSITE: PS00464; RIBOSOMAL_L22; UNKNOWN 1.
SQ SEQUENCE 1498 AA; 167705 MW; AADEA3FFD1ABFEB CRC64;

Query Match      42.9%; Score 6; DB 3; Length 1498;
Best Local Similarity 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 0;

QY 6 PYEGR 11
DB 436 PYEGR 441

RESULT 10
O18923 O18923 PRELIMINARY; PRT; 40 AA.
AC O18923;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CADHERIN-11 (OSTEOBLAST-CADHERIN) (OSF-4) (FRAGMENT).
GN CH11.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-RETINA;
RA Lutz D.A., Zheng J.J.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC

```

CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.

DR EMBL: AF033020; AAB87475.1; -.
 DR InterPro: IPR002126; Cadherin.
 KW PROSITE; PS00232; CADHERIN_1; PARTIAL.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding.
 FT NON_TER 1
 FT CHAIN <1
 FT NON_TER 40
 FT NON_TER 40
 SQ SEQUENCE 40 AA; 4373 MW; B554B44C802BDAFD CRC64;

Query Match 35.7%; Score 5; DB 6; Length 40;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYEGR 11
 DB 6 GYEGR 10

RESULT 11
 Q9PWNO Q9PWNO PRELIMINARY; PRT; 40 AA.
 AC Q9PWNO;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE CADHERIN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Askelevich F., Zhu Y., Ranscht B., Sanes J.R.;
 RT "Expression of Multiple Cadherins and Catenins in the Chick Optic
 RL Tectum".
 RL Mol. Cell. Neurosci. 0:0-0(1999).
 DR EMBL: AF098469; AAD16259.1; -.
 FT NON_TER 1
 FT NON_TER 40
 FT NON_TER 40
 SQ SEQUENCE 40 AA; 4373 MW; B554B44C802BDAFD CRC64;

Query Match 35.7%; Score 5; DB 13; Length 40;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYEGR 11
 DB 6 GYEGR 10

RESULT 12
 O18958 O18958 PRELIMINARY; PRT; 42 AA.
 AC O18958;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE BONE MORPHOGENETIC PROTEIN 1 (FRAGMENT).
 GN BMP1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-98266882; PubMed-9605845;
 RA Martin-Burriel I., Goldammer T., Eiduque C., Lunidin M., Barendse W.,
 Zaragoza P., Olsaker I.;
 RT "Physical and linkage mapping of the bovine bone morphogenetic protein
 1 on the evolutionary break region of BTA 8.";
 RL Cytogenet. Cell Genet. 79:179-183(1997).
 DR EMBL; Y14605; CAA74948.1; -;
 DR HSP; P00736; IAPQ.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR Pfam; PF00008; EGF; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 FT NON_TER 1
 FT NON_TER 42
 SQ SEQUENCE 42 AA; 4739 MW; 4E5967160BCF9B24 CRC64;

Query Match 35.7%; Score 5; DB 6; Length 42;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DPGYE 9
 Db 29 DPGYE 33

RESULT 13
 ID 004128 PRELIMINARY; PRT; 69 AA.
 AC 004128;
 DT 01-JUL-1997 (TRENBLrel. 04, Created)
 DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
 DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 8.0 KDA PROTEIN (FRAGMENT).
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B73 INBRED LINE;
 RX MEDLINE-97248483; PubMed-9094711;
 RA Gray J., Close P.S., Briggs S.P., Johal G.S.;
 RT "A novel suppressor of cell death in plants encoded by the Lls1 gene
 of maize.";
 RL Cell 89:25-31(1997).
 DR EMBL; U77346; AAC49677.1; -;
 DR Mendel; 23979; Zeama; 2476; 23979.
 DR InterPro; IPR001395; Aldo_ket_red.
 DR Pfam; PF00248; aldo_ket_red; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 69
 SQ SEQUENCE 69 AA; 7954 MW; 2F4976D13CAD36CF CRC64;

Query Match 35.7%; Score 5; DB 10; Length 69;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GRPEY 14
 Db 30 GRPEY 34

RESULT 14
 Q9QVS3

ID Q9QVS3 PRELIMINARY; PRT; 101 AA.
 AC Q9QVS3;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE TISSUE TYPE II TRANSGUTAMINASE.
 OS Rattus sp.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10118;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95251649; PubMed-7733921;
 RA Iwai N., Shimoike H., Kinoshita M.;
 RT "Genes up-regulated in hypertrophied ventricle.";
 RL Biochem. Biophys. Res. Commun. 209:527-534(1995).
 DR HSP; P00488; IF13.
 SQ SEQUENCE 101 AA; 11214 MW; D1ECD4548E11469 CRC64;

Query Match 35.7%; Score 5; DB 11; Length 101;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGYEG 10
 Db 14 PGYEG 18

RESULT 15
 ID Q9JIW2 PRELIMINARY; PRT; 106 AA.
 AC Q9JIW2;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE CADHERIN-11 (FRAGMENT).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISCHER344; TISSUE=TESTIS;
 RX MEDLINE-20114371; PubMed-10650949;
 RA Johnson K.J., Patel S.R., Boekelheide K.;
 RT "Multiple cadherin superfamily members with unique expression profiles
 are produced in rat testis.";
 RL Endocrinology 141:675-683(2000).
 DR EMBL; AF177677; AAF87052.1; -;
 DR InterPro; IPR000233; Cadherin_C_term.
 DR Pfam; PF01049; Cadherin_C_term; 1.
 FT NON_TER 1
 FT NON_TER 106
 SQ SEQUENCE 106 AA; 11775 MW; 3F633A8793AB1B9F CRC64;

Query Match 35.7%; Score 5; DB 11; Length 106;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYEGR 11
 Db 72 GYEGR 76

Search completed: January 31, 2002, 13:37:43
 Job time: 168 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:39:19 ; Search time 46.78 Seconds
(without alignments)
10.973 Million cell updates/sec

Title: US-08-957-709-75

Perfect score: 14

Sequence: 1 SAVHDPGEGRPPEY 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	50.0	1196	1 DNBI_HSV2H	P9452 herpes simp
2	7	50.0	1197	1 DNBI_HSV2	P36384 herpes simp
3	6	42.9	1025	1 CA16_MOUSE	Q04857 mus musculu
4	5	35.7	88	1 IE12_HSV11	P03170 herpes simp
5	5	35.7	119	1 RL20_XYLFA	Q9pf08 xylella fas
6	5	35.7	148	1 NEU2_CATCO	P15211 catostomus
7	5	35.7	154	1 NEU1_CATCO	P15210 catostomus
8	5	35.7	208	1 YJ79_AQUAE	O67788 aquifex aeo
9	5	35.7	219	1 RPIA_ECOLI	P27252 escherichia
10	5	35.7	252	1 GSC_HUMAN	P56915 homo sapien
11	5	35.7	256	1 GSC_MOUSE	Q02591 mus musculu
12	5	35.7	272	1 IL2A_HUMAN	P01589 homo sapien
13	5	35.7	281	1 RS2_CHLMU	P71145 chlamydia m
14	5	35.7	282	1 RS2_CHLTR	O84687 chlamydia t
15	5	35.7	314	1 IUNH_CRIFA	Q27546 crithidia f
16	5	35.7	341	1 SPIN_CBEPU	P23061 christstoneu
17	5	35.7	341	1 Y422_TREPA	O83437 treponema p
18	5	35.7	344	1 ARC1_THETH	P96136 thermus aqu
19	5	35.7	349	1 RPA1_HALMO	P15349 halococcus
20	5	35.7	351	1 SPIN_HAEPU	Q05894 heliothis a
21	5	35.7	373	1 VU3_HSV6U	Q01350 human herpe
22	5	35.7	396	1 PRTZ_BOVIN	P00744 bos taurus
23	5	35.7	397	1 TDG_MOUSE	P56581 mus musculu
24	5	35.7	400	1 PRTZ_HUMAN	P22891 homo sapien
25	5	35.7	408	1 BPHG_BURCE	P37337 burkholderi
26	5	35.7	409	1 PROA_MYCLE	Q9cbz7 mycobacteri
27	5	35.7	410	1 BEDA_PSEPU	Q07946 pseudomonas
28	5	35.7	410	1 TDG_HUMAN	Q13569 homo sapien
29	5	35.7	415	1 PROA_MYCTU	P71921 mycobacteri
30	5	35.7	419	1 CBP1_HUMAN	P15085 homo sapien
31	5	35.7	419	1 CBPA_BOVIN	P00730 bos taurus
32	5	35.7	419	1 PROA_STRCO	Q9rdk1 streptomyce
33	5	35.7	423	1 YWV2_CABEL	Q11076 caenorhabdi

34 5 35.7 437 1 ERR1_YEAST P42222 saccharomyc
35 5 35.7 448 1 FBL5_HUMAN Q9ubx5 homo sapien
36 5 35.7 448 1 FBL5_MOUSE Q9vvh9 mus musculu
37 5 35.7 448 1 FBL5_RAT Q9vvh8 rattus norv
38 5 35.7 454 1 VNUC_THOGV P89216 thogoto vir
39 5 35.7 462 1 REGB_RHOSH Q53068 rhodobacter
40 5 35.7 475 1 MOTD_RHIME Q52964 rhizobium m
41 5 35.7 504 1 YC03_KLEPN Q48449 klebsiella
42 5 35.7 535 1 ARS_PSEAE P51691 pseudomonas
43 5 35.7 547 1 RCCL_DROME P25171 drosophila
44 5 35.7 552 1 PFPB_RICCO Q41141 ricinus com
45 5 35.7 552 1 PFPB_SOLTU P21343 solanum tub

ALIGNMENTS

RESULT 1

ID DNBI_HSV2H STANDARD: PRT: 1196 AA.
AC P89452:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR DNA-BINDING PROTEIN (INFECTED CELL PROTEIN 8) (ICP 8 PROTEIN).
GN DBP OR UL29 OR ICP8.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA REPLICATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 286099; CAB06754.1; -
CC InterPro: IPR000635; Viral_DNA_bind.
CC Pfam; PF00747; viral_DNA_bp; 1.
CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
CC ZN_FING 499 512 C2HC-TYPE.
CC FT DOMAIN 1168 1196 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC SQ SEQUENCE 1196 AA; 128423 MW; A19CA843280DD7F5 CRC64;
CC -----

Query Match 50.0%; Score 7; DB 1; Length 1196;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGYEGRP 12
DB 429 PGYEGRP 435

RESULT 2

ID DNBI_HSV2 STANDARD: PRT: 1197 AA.
AC P89452;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MAJOR DNA-BINDING PROTEIN (INFECTED CELL PROTEIN 8) (ICP 8 PROTEIN).
GN DBP OR UL29 OR ICP8.
OS Herpes simplex virus (type 2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OX Alphaherpesvirinae; Simplexvirus.
NCBI_TaxID=10310;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=93228441; PubMed=8385914;
RA Toh Y., Liu Y., Tanaka S., Mori R.;
RT "Nucleotide sequence of the major DNA-binding protein gene of herpes
RT simplex virus type 2 and a comparison with the type 1.";
RL Arch. Virol. 129:183-196(1993).
CC 1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC 1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC 1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
DR PIR: A48350;
DR InterPro: IPR000635; Viral DNA bind.
DR Pfam: PF00747; viral DNA bp; 1
KW DNA-binding; DNA recognition; Zinc-finger; Nuclear protein.
FT 2N-FING 499 512 C2HG-TYPE
FT DOMAIN 1169 1197 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 1197 AA; 128412 MW; C1576B5B865BFB CRC64;

Query Match 50.0%; Score 7; DB 1; Length 1197;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 6 PGYGRP 12
Db 429 PGYGRP 435
|||||||

RESULT 3
ID CA16_MOUSE STANDARD; PRT; 1025 AA.
AC Q04857;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(VI) CHAIN PRECURSOR.
GN COL6A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=93316904; PubMed=8326912;
RA Bonaldo P., Piccolo S., Marvulli D., Volpin D., Bressan G.M.;
RT "Murine alpha 1(VI) collagen chain. Complete amino acid sequence and
RT identification of the gene promoter region.";
RL Matrix 13:223-233(1993).
[2]
RN SEQUENCE OF 442-1025 FROM N.A.
RX MEDLINE=93256886; PubMed=8489506;
RA Zhang R.Z., Pan T.C., Timpl R., Chu M.L.;
RT "Cloning and sequence analysis of cDNAs encoding the alpha 1, alpha 2
RT and alpha 3 chains of mouse collagen VI.";
RL Biochem. J. 291:787-792(1993).
CC 1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN
CC ALPHA 2(VI) AND ALPHA 3(VI).
CC 1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),
CC ALPHA 2(VI) AND ALPHA 3(VI).
CC 1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC 1- SIMILARITY: CONTAINS 3 VWFA DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; X66405; CAA47032.1; -;
CC EMBL; X66406; CAA47033.1; -;
CC EMBL; Z18271; CAA79152.1; -;
CC PIR: S34839; S34839.
CC MGD: MGI-88459; Col6a1.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR002035; VWFA.
CC Pfam: PF01391; Collagen; 5.
CC Pfam: PF00092; vwa; 3
CC PRINTS: PR00453; VWFADOMAIN.
CC SMART: SM00327; VWA; 3
CC PROSITE: PS50234; VWFA; 3
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Cell adhesion; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1025
FT DOMAIN 20 255
FT DOMAIN 256 591
FT DOMAIN 592 1025
FT DOMAIN 36 234
FT DOMAIN 614 802
FT DOMAIN 826 1018
FT SITE 261 263
FT SITE 441 443
FT SITE 477 479
FT SITE 511 511
FT CARBOHYD 515 515
FT CARBOHYD 536 536
FT CARBOHYD 801 801
FT CARBOHYD 893 893
FT CONFLICT 674 675
FT CONFLICT 709 709
FT CONFLICT 943 943
FT CONFLICT 960 960
SQ SEQUENCE 1025 AA; 108489 MW; 2A05DFED8771BBF7 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 1025;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 DPGYEG 10
Db 263 DPGYEG 268
|||||||

RESULT 4
ID IE12_HSV11 STANDARD; PRT; 88 AA.
AC P03170;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE IMMEDIATE-EARLY PROTEIN IE12 (IMMEDIATE-EARLY-5) (VMW12).
GN US12.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNaab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574(1988).

RN [2] SEQUENCE FROM N.A.
 RP MEDLINE=82150256; PubMed=6278443;
 RX Watson R.J., Vande Woude G.F.;
 RA "DNA sequence of an immediate-early gene (IE-mRNA-5) of herpes simplex
 RT virus type 1.";
 RL Nucleic Acids Res. 10:979-991(1982).
 RN [3]
 RP SEQUENCE OF 1-55 FROM N.A.
 RX MEDLINE=82059456; PubMed=6272198;
 RA Watson R.J., Umene K., Enquist L.W.;
 RT "Reiterated sequences within the intron of an immediate-early gene of
 RL herpes simplex virus type 1.";
 RN Nucleic Acids Res. 9:4189-4199(1981).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83032477; PubMed=6290591;
 RA Murchie M.-J., McGeoch D.J.;
 RT "DNA sequence analysis of an immediate-early gene region of the
 RL herpes simplex virus type 1 genome (map coordinates 0.950 to
 J. Gen. Virol. 62:1-15(1982).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85160822; PubMed=2984429;
 RA McGeoch D.J., Dolan A., Donald S., Rixon F.J.;
 RT "Sequence determination and genetic content of the short unique
 RL region in the genome of herpes simplex virus type 1.";
 RN J. Mol. Biol. 181:1-13(1985).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84169548; PubMed=6324121;
 RA Rixon F.J., McGeoch D.J.;
 RT "A 3' co-terminal family of mRNAs from the herpes simplex virus type
 RL 1 short region: two overlapping reading frames encode unrelated
 polypeptide one of which has highly reiterated amino acid sequence.";
 RN Nucleic Acids Res. 12:2473-2487(1984).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J02220; AAA45795.1; -
 DR EMBL: L00036; AAA96676.1; -
 DR EMBL: X14112; CAA32277.1; -
 DR EMBL: V00462; CAA23737.1; -
 DR EMBL: X00428; CAA25124.1; -
 DR EMBL: X02138; CAA26066.1; -
 DR PIR: A03724; EDBE51.
 KW Early protein.
 FT CONFLICT 14 14 N -> T (IN REF. 4, 5 AND 6).
 SQ SEQUENCE 88 AA; 9806 MW; C5D2F662C71C88E0 CRC64;

 Query Match 35.7%; Score 5; DB 1; Length 88;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 2 AVHDP 6
 Db 43 AVHDP 47
 |||||

 RESULT 5
 RL20_XYLFA STANDARD; PRT; 119 AA.
 AC Q9PFB8;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 50S RIBOSOMAL PROTEIN L20.
 GN RPLT OR XF0740.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 CC -----
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
 CC NECESSARY TO THE IN VITRO ASSEMBLY PROCESS OF THE 50S RIBOSOMAL
 CC SUBUNIT. IT IS NOT INVOLVED IN THE PROTEIN SYNTHESIZING FUNCTIONS
 CC OF THAT SUBUNIT (BY SIMILARITY).
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE003916; AAF83550.1; -
 DR InterPro: IPR001081; Ribosomal_L20.
 DR Pfam: PF00453; Ribosomal_L20; 1.
 DR PRINTS: PR00062; RIBOSOMALL20.
 DR ProDom: PD002389; Ribosomal_L20; 1.
 DR PROSITE: PS00937; RIBOSOMAL_L20; 1.
 KW Ribosomal protein; rRNA-binding; Complete proteome.
 SQ SEQUENCE 119 AA; 13499 MW; FC2C9C68D3C534BB CRC64;

 Query Match 35.7%; Score 5; DB 1; Length 119;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 2 AVHDP 6
 Db 99 AVHDP 103
 |||||

 RESULT 6

```

NEU2_CATCO STANDARD; PRT; 148 AA.
ID NEU2_CATCO
AC P15211;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ISOTOCIN-NEUROPHYSIN IT 2 PRECURSOR [CONTAINS: ISOTOCIN (IT)];
DE NEUROPHYSIN IT 2].
OS Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Catostomidae; Catostomus.
OX NCBI_TaxID=7971;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90059870; PubMed=2583084;
RA Figueroa J., Morley S.D., Heierhorst J., Krentler C., Lederis K.,
RA Richter D.;
RT "Two isotocin genes are present in the white sucker Catostomus
commersoni both lacking introns in their protein coding regions.";
RL EMBO J. 8:2873-2877(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315773; PubMed=2748582;
RA Heierhorst J., Morley S.D., Figueroa J., Krentler C., Lederis K.,
RA Richter D.;
RT "Vasotocin and isotocin precursors from the white sucker, Catostomus
commersoni: cloning and sequence analysis of the cDNAs.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:5242-5246(1989).
CC -1- FUNCTION: ISOTOCIN CAUSES CONTRACTION OF SMOOTH MUSCLES.
CC -1- PTM: SEVEN DISULFIDE BONDS ARE PRESENT IN NEUROPHYSIN.
CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16622; CAA34619.1;
DR PIR: S06019; S06019.
DR HSSP: P01180; INPO.
DR InterPro: IPR000981; Neurohypophys_horm.
DR Pfam: PF00220; hormone4; 1.
DR Pfam: PF00184; hormone5; 1.
DR PRINTS: PR00831; NEUROPHYSIN.
DR ProDom: PD001676; Neurohypophys_horm; 1.
DR SMART: SM00003; NH; 1.
DR PROSITE: PS00284; NEUROHYPOPHYS_HORM; 1.
DR Hormone; Hypothalamus; Amidation; Cleavage on pair of basic residues;
KW Multigene family; Signal.
FT SIGNAL 1 20
FT PEPTIDE 21 29 ISOTOCIN.
FT PEPTIDE 32 148 NEUROPHYSIN IT 2.
FT DISULFID 21 26
FT MOD_RES 29 29
SQ SEQUENCE 148 AA; 15581 MW; 26A9EDC2513E9A6F CRC64;

Query Match 35.7%; Score 5; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYGR 11
Db 94 GYGR 98

RESULT 7
NEU1_CATCO STANDARD; PRT; 154 AA.
ID NEU1_CATCO
AC P15210;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ISOTOCIN-NEUROPHYSIN IT 1 PRECURSOR [CONTAINS: ISOTOCIN (IT)];
DE NEUROPHYSIN IT 1].
OS Catostomus commersoni (White sucker).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Catostomidae; Catostomus.
OX NCBI_TaxID=7971;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90059870; PubMed=2583084;
RA Figueroa J., Morley S.D., Heierhorst J., Krentler C., Lederis K.,
RA Richter D.;
RT "Two isotocin genes are present in the white sucker Catostomus
commersoni both lacking introns in their protein coding regions.";
RL EMBO J. 8:2873-2877(1989).
CC -1- FUNCTION: ISOTOCIN CAUSES CONTRACTION OF SMOOTH MUSCLES.
CC -1- PTM: SEVEN DISULFIDE BONDS ARE PRESENT IN NEUROPHYSIN.
CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16622; CAA34619.1;
DR PIR: S06019; S06019.
DR HSSP: P01180; INPO.
DR InterPro: IPR000981; Neurohypophys_horm.
DR Pfam: PF00220; hormone4; 1.
DR Pfam: PF00184; hormone5; 1.
DR PRINTS: PR00831; NEUROPHYSIN.
DR ProDom: PD001676; Neurohypophys_horm; 1.
DR SMART: SM00003; NH; 1.
DR PROSITE: PS00284; NEUROHYPOPHYS_HORM; 1.
DR Hormone; Hypothalamus; Amidation; Cleavage on pair of basic residues;
KW Multigene family; Signal.
FT SIGNAL 1 20
FT PEPTIDE 21 29 ISOTOCIN.
FT PEPTIDE 32 148 NEUROPHYSIN IT 2.
FT DISULFID 21 26
FT MOD_RES 29 29
SQ SEQUENCE 148 AA; 15581 MW; 26A9EDC2513E9A6F CRC64;

Query Match 35.7%; Score 5; DB 1; Length 148;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYGR 11
Db 94 GYGR 98

RESULT 8
YJ79_AQUAE STANDARD; PRT; 208 AA.
ID YJ79_AQUAE
AC Q67788;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL ALDOLASE CLASS II PROTEIN AQ_1979.
DE AQ_1979.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=VF5;
RX  MEDLINE=98196666; PubMed=9537320;
RA  Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA  Graham D.E., Overbeek R., Sneed M.A., Keller M., AuJay M., Huber R.,
RA  Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT  "The complete genome of the hyperthermophilic bacterium Aquifex
RT  aeolicus.";
RL  Nature 392:353-358(1998).
CC  -1- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (POTENTIAL).
CC  -1- SIMILARITY: BELONGS TO THE ALDOLASE CLASS II FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: AE000766; AAC07751.1; -.
DR  InterPro: IPR001303; Aldolase_II.
DR  Pfam: PF00596; Aldolase_II; 1.
KW  Hypothetical protein; Zinc; Complete proteome.
FT  METAL 80 80 ZINC (BY SIMILARITY).
FT  METAL 99 99 ZINC (BY SIMILARITY).
FT  METAL 101 101 ZINC (BY SIMILARITY).
FT  METAL 174 174 ZINC (BY SIMILARITY).
SQ  SEQUENCE 208 AA; 23550 MW; 3295652C5ED17344 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 208;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 YEGRP 12
Db 67 YEGRP 71

RESULT 9
RPIA_ECOLI STANDARD; PRT; 219 AA.
AC P27252;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RPIA OR B2914 OR Z4252 OR ECS3785.
GN Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN  SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RP  STRAIN=K12;
RX  MEDLINE=93374859; PubMed=8366047;
RA  Hove-Jensen B., Maigaard M.;
RT  "Escherichia coli rpiA gene encoding ribose phosphate isomerase A.";
RL  J. Bacteriol. 175:5628-5635(1993).
RN  [2]
RP  SEQUENCE FROM N.A.
RP  STRAIN=K12;
RA  Roy I., Leadlay P.F.;
RL  Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RX  STRAIN=K12 / MGL655;
RX  MEDLINE=97426617; PubMed=9278503;
RA  Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```

```

RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12.";
RL  Science 277:1453-1474(1997).
RN  [4]
RP  SEQUENCE FROM N.A.
RP  STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX  MEDLINE=21074935; PubMed=11206551;
RA  Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA  Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA  Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA  Welch R.A., Blattner F.R.;
RT  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL  Nature 409:529-533(2001).
RN  [5]
RP  SEQUENCE FROM N.A.
RP  STRAIN=O157:H7 / RIMD 0509952;
RX  MEDLINE=21156231; PubMed=11258796;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
RN  [6]
RP  SEQUENCE OF 154-219 FROM N.A.
RP  STRAIN=K12;
RX  MEDLINE=92011350; PubMed=1917830;
RA  Rex J.H., Aronson B.D., Somerville R.L.;
RT  "The tdh and serA operons of Escherichia coli: mutational analysis of
RT  the regulatory elements of leucine-responsive genes.";
RL  J. Bacteriol. 173:5944-5953(1991).
RN  [7]
RP  SEQUENCE OF 1-12.
RP  STRAIN=K12 / EMG2;
RX  MEDLINE=97443975; PubMed=9298646;
RA  Link A.J., Robison K., Church G.M.;
RT  "Comparing the predicted and observed properties of proteins encoded
RT  in the genome of Escherichia coli K-12.";
RL  Electrophoresis 18:1259-1313(1997).
CC  -1- CATALYTIC ACTIVITY: D-RIBOSE 5-PHOSPHATE -> D-RIBULOSE 5-PHOSPHATE.
CC  -1- PATHWAY: NONOXIDATIVE BRANCH OF THE PENTOSE PHOSPHATE PATHWAY.
CC  -1- SUBUNIT: HOMODIMER.
CC  -1- SIMILARITY: BELONGS TO THE RIBOSE 5-PHOSPHATE ISOMERASE FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: X73026; CAA51509.1; -.
DR  EMBL: X66836; CAA47309.1; ALT_INIT.
DR  EMBL: U28377; AAA69081.1; -.
DR  EMBL: AE000374; AAC75951.1; -.
DR  EMBL: AE005521; AAG58041.1; -.
DR  EMBL: AP002563; BAB37208.1; -.
DR  EMBL: M64630; AAA73015.1; -.
DR  PIR: S22097; S22097.
DR  PIR: S37500; S37500.
DR  PIR: A48645; A48645.
DR  SWISS-2DPAGE; P27252; COLI.
DR  EcoGene; EG11443; rpiA.
KW  Isomerase; Complete proteome.
SQ  SEQUENCE 219 AA; 22860 MW; B53C49CC3DB188BC CRC64;

Query Match 35.7%; Score 5; DB 1; Length 219;

```

Best Local Similarity 100.0%; Pred. No. 44; Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 GRPEY 14
Db 152 GRPEY 156

RESULT 10
GSC_HUMAN STANDARD; PRT; 252 AA.

AC P56915;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HOMEBOX PROTEIN GOOSECOID.
GN GSC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte.
RX MEDLINE=94375063; PubMed=7916327;
RA Blum M., De Robertis E.M., Kojis T., Heinzmann C., Klisak I.,
RA Geissert D., Sparkes R.S.,
RT "Molecular cloning of the human homeobox gene goosecoid (GSC) and
FL mapping of the gene to human chromosome 14q32.1.";
RL Genomics 21:388-393 (1994)

CC -1- FUNCTION: REGULATES CHORDIN (CHRD). MAY PLAY A ROLE IN SPATIAL
CC COMPARTMENTS DURING ORGANOGENESIS.
CC PROGRAMING WITHIN DISCRETE EMBRYONIC FIELDS OR LINEAGE
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.

CC "BICOID" SUBFAMILY.
DR HSP; P06601; IFLJL.
DR MIM; I38890.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Developmental protein; Nuclear protein; DNA-binding; Homeobox.
FT DNA_BIND 155 214 HOMEBOX.
SQ SEQUENCE 252 AA; 27853 MW; 6C0EEC48C084D323 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGYEG 10
Db 113 PGYEG 117

RESULT 11
GSC_MOUSE STANDARD; PRT; 256 AA.

AC Q02591;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMEBOX PROTEIN GOOSECOID.
GN GSC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315328; PubMed=1352187;

RA Blum M., Gaunt S.J., Cho K.W.Y., Steinbeisser H., Blumberg B.,
RA Bittner D.A., de Robertis E.M.;
RT "Gastrulation in the mouse: the role of the homeobox gene goosecoid.";
RL Cell 69:1097-1106(1992).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=98079105; PubMed=9417125;
RA Danilov V., Blum M., Schweickert A., Campione M., Steinbeisser H.;
RT "Negative autoregulation of the organizer-specific homeobox gene
goosecoid.";
RL J. Biol. Chem. 273:627-635(1998).

CC -1- FUNCTION: GOOSECOID-EXPRESSING REGIONS OF THE GASTRULATING
CC MOUSE EGG CYLINDER HAVE ORGANIZER-LIKE ACTIVITY WHEN TRANSPLANTED
CC INTO XENOPUS EMBRYOS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: IN EARLY GASTRULATION, EXPRESSED IN THE
CC DORSAL LIP. IN LATER STAGES OF DEVELOPMENT FOUND IN HEAD, LIMBS
CC AND BODY WALL.
CC -1- INDUCTION: BY ACTIVIN.
CC -1- SIMILARITY: BELONGS TO THE PAIRED FAMILY OF HOMEBOX PROTEINS.

CC "BICOID" SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M85271; AAA37826.1;
DR EMBL; Y13145; CAA73611.1;
DR EMBL; Y13150; CAA73612.1;
DR PIR; A42768; A42768.
DR HSP; P06601; IFLJL.
DR TRANSFAC; T02047;
DR MGD; MGI:95841; Gsc.
DR InterPro: IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Developmental protein; Nuclear protein; DNA-binding; Homeobox.
FT DNA_BIND 160 219 HOMEBOX.
SQ SEQUENCE 256 AA; 27979 MW; 3639FB059AC3DB9E CRC64;

Query Match 35.7%; Score 5; DB 1; Length 256;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGYEG 10
Db 118 PGYEG 122

RESULT 12
IL2A_HUMAN STANDARD; PRT; 272 AA.

AC P01589;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTERLEUKIN-2 RECEPTOR ALPHA CHAIN PRECURSOR (IL-2 RECEPTOR ALPHA
DE SUBUNIT) (P55) (TAC ANTIGEN) (CD25 ANTIGEN).
GN IL2RA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=85012734; PubMed=6090949;
RA Nikaïdo T., Shimizu A., Ishida N., Sabe H., Teshigawara K., Maeda M.,
RA Uchiyama T., Yodoi J., Honjo T.;
RT "Molecular cloning of cDNA encoding human interleukin-2 receptor.";
RL Nature 311:631-635(1984).
[2]
RX SEQUENCE FROM N.A.
RX MEDLINE=85012733; PubMed=6090948;
RA Leonard W.J., Depper J.M., Crabtree G.R., Rudikoff S., Pumphrey J.,
RA Robb R.J., Kroenke M., Svetlik P.B., Peffer N.J., Waldmann T.A.,
RA Greene W.C.;
RT "Molecular cloning and expression of cDNAs for the human
RT interleukin-2 receptor.";
RL Nature 311:626-631(1984).
[3]
RX SEQUENCE FROM N.A.
RX MEDLINE=86067183; PubMed=2999698;
RA Ishida N., Kanamori H., Noma T., Nikaïdo T., Sabe H., Suzuki N.,
RA Shimizu A., Honjo T.;
RT "Molecular cloning and structure of the human interleukin 2 receptor
RT gene.";
RL Nucleic Acids Res. 13:7579-7589(1985).
[4]
RX SEQUENCE FROM N.A.
RX MEDLINE=86018868; PubMed=2996141;
RA Leonard W.J., Depper J.M., Kanehisa M., Kroenke M., Peffer N.J.,
RA Svetlik P.B., Sullivan M., Greene W.C.;
RT "Structure of the human interleukin-2 receptor gene.";
RL Science 230:633-639(1985).
[5]
RX SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=87159546; PubMed=3030566;
RA Cross S.L., Feinberg M.B., Wolf J.B., Holbrook N.J., Wong-Stall F.,
RA Leonard W.J.;
RT "Regulation of the human interleukin-2 receptor alpha chain promoter:
RT activation of a nonfunctional promoter by the transactivator gene of
RT HTLV-I";
RL Cell 49:47-56(1987).
[6]
RX 3D-STRUCTURE MODELING OF 23-83.
RX MEDLINE=95111955; PubMed=7529123;
RA Bamorough P., Hedgecock C.J., Richards W.G.;
RT "The interleukin-2 and interleukin-4 receptors studied by molecular
RT modelling.";
RL Structure 2:839-851(1994).
CC -1- FUNCTION: RECEPTOR FOR INTERLEUKIN-2.
CC -1- SUBUNIT: NON COVALENT DIMER OF AN ALPHA AND A BETA CHAINS. IL-2R
CC EXIST IN 3 DIFFERENT FORMS: A HIGH AFFINITY DIMER, AN INTERMEDIATE
CC AFFINITY MONOMER (BETA CHAIN), AND A LOW AFFINITY MONOMER (ALPHA
CC CHAIN). THE HIGH AND INTERMEDIATE AFFINITY FORMS ALSO ASSOCIATE
CC WITH A GAMMA CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD25 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd25.htm".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X01057; CAA25525.1; -
CC EMBL; X01313; CAA26906.1; -
CC EMBL; X03132; CAA26906.1; JOINED.
CC EMBL; X03133; CAA26906.1; JOINED.
CC EMBL; X03134; CAA26906.1; JOINED.
CC EMBL; X03135; CAA26906.1; JOINED.
CC EMBL; X03136; CAA26906.1; JOINED.
CC EMBL; X03137; CAA26906.1; JOINED.
CC EMBL; X03138; CAA26906.1; JOINED.

DR EMBL; K03122; AAB59535.1; ALT_SEQ.
DR EMBL; M11066; AAA67527.1; -
DR EMBL; M10322; AAA67527.1; JOINED.
DR EMBL; M11060; AAA67527.1; JOINED.
DR EMBL; M11061; AAA67527.1; JOINED.
DR EMBL; M11062; AAA67527.1; JOINED.
DR EMBL; M11063; AAA67527.1; JOINED.
DR EMBL; M11064; AAA67527.1; JOINED.
DR EMBL; M11065; AAA67527.1; JOINED.
DR EMBL; M15864; AAA59162.1; -
DR PIR; A01856; UHHU2.
DR PIR; A24113; A24113.
DR PIR; A44186; A44186.
DR PDB; 1ILM; 26-JAN-95.
DR PDB; 1ILN; 26-JAN-95.
DR MIM; 147730; -
DR InterPro; IPR000436; Sush1_SCR_CCP.
DR Pfam; PF00084; sush1; 2.
DR SMART; SM00032; CCP; 2.
KW Transmembrane; Glycoprotein; Receptor; Repeat; Signal; Sushi;
KW 3D-structure. 1 21
FT SIGNAL 1 21 INTERLEUKIN-2 RECEPTOR ALPHA CHAIN.
FT CHAIN 22 272 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 22 240 POTENTIAL.
FT TRANSMEM 241 259 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 260 272 SUSHI 1.
FT DOMAIN 23 81 SUSHI 2.
FT DOMAIN 124 185 BY SIMILARITY.
FT DISULFID 24 67 BY SIMILARITY.
FT DISULFID 751 80 BY SIMILARITY.
FT DISULFID 125 168 BY SIMILARITY.
FT DISULFID 152 184 BY SIMILARITY.
FT CARBOHYD 70 70 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 272 AA; 30819 MW; 83D907C8C81D2C0E CRC64;

Query Match 35.7%; Score 5; DB 1; Length 272;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 EGRPE 13

DB 204 EGRPE 208

RESULT 13

ID RS2_CHLMU STANDARD; PRT; 281 AA.
AC P71145;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S2.
GN RPSB OR TC0051

OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=83560;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MOPN;

RX MEDLINE=97386531; PubMed=9244380;

RA Zhang Y.X., Tao J., Zhou M., Meng Q., Zhang L., Shen L., Klein R.,

RA Miller D.L.;

RT "Elongation factor Ts of Chlamydia trachomatis: structure of the gene
RT and properties of the protein";

RL Arch. Biochem. Biophys. 344:43-52(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=MOPN / NIGG;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K.,

RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
RA Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U60196; AAB07069.1; -;
DR EMBL: AE002272; AAF38940.1; -;
DR TIGR: TC0051; -;
DR InterPro: IPR001865; Ribosomal_S2.
DR Pfam: PF00318; Ribosomal_S2; 1.
DR PRINTS: PR00395; RIBOSOMAL_S2.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; 1.
DR PROSITE: PS00963; RIBOSOMAL_S2_2; 1.
DR Ribosomal protein; Complete proteome.
KW Ribosomal protein; Complete proteome.
FT CONFLICT 176 176 A--R (IN REF. 1).
SQ SEQUENCE 281 AA; 31079 MW; 29E4F309F45435F7 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 281;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 DPGYE 9
Db 167 DPGYE 171

RESULT 14
RS2_CHLTR STANDARD; PRT; 282 AA.
AC 084687; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S2.
GN RPSB OR RS2 OR CT680.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Ollinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001338; AAC68275.1; -;
DR InterPro: IPR001865; Ribosomal_S2.
DR Pfam: PF00318; Ribosomal_S2; 1.

DR PRINTS: PR00395; RIBOSOMALS2.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; 1.
DR PROSITE: PS00963; RIBOSOMAL_S2_2; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 282 AA; 31146 MW; 63693C13CA6EA26B CRC64;

Query Match 35.7%; Score 5; DB 1; Length 282;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 DPGYE 9
Db 167 DPGYE 171

RESULT 15
IUNH_CRIFA STANDARD; PRT; 314 AA.
AC 027546;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE INOSINE-URIDINE PREFERING NUCLEOSIDE HYDROLASE (EC 3.2.2.1)
DE (IU-NUCLEOSIDE HYDROLASE) (PURINE NUCLEOSIDASE).
GN IUNH
OS Crithidia fasciculata.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
OX NCBI_TaxID=5636;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND MUTAGENESIS.
RX MEDLINE=96214502; PubMed=8634237;
RA Gopaul D.N., Meyer S.L., Degano M., Sacchetti J.C., Schramm V.L.;
RT "Inosine-uridine nucleoside hydrolase from Crithidia fasciculata.
RT Genetic characterization, crystallization, and identification of
RT histidine 241 as a catalytic site residue.";
RL Biochemistry 35:5963-5970(1996).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96214503; PubMed=8634238;
RA Degano M., Gopaul D.N., Scapin G., Schramm V.L., Sacchetti J.C.;
RT "Three-dimensional structure of the inosine-uridine nucleoside N-
RT ribohydrolase from Crithidia fasciculata.";
RL Biochemistry 35:5971-5981(1996).
CC -1- FUNCTION: CATALYZES THE HYDROLYSIS OF ALL OF THE COMMONLY
CC OCCURRING PURINE AND PYRIMIDINE NUCLEOSIDES INTO RIBOSE AND THE
CC ASSOCIATED BASE, BUT HAS A PREFERENCE FOR INOSINE AND URIDINE AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: A N-D-RIBOSYLPURINE + H(2)O = A PURINE + D-
CC RIBOSE.
CC -1- PATHWAY: PURINE SALVAGE.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- MASS SPECTROMETRY: MW=34194; MW_ERR=4; METHOD=ELECTROSPRAY.
CC -1- SIMILARITY: BELONGS TO THE IUNH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U43371; AAC47119.1; -;
DR PDB: 1MAS; 17-AUG-96.
DR PDB: 2MAS; 12-AUG-97.
DR InterPro: IPR001910; IUNH.
DR Pfam: PF01136; IU_nuc_hydro; 1.
DR ProDom: PD007736; IUNH; 1.
DR PROSITE: PS01247; IUNH; 1.
DR Hydrolase; 3D-structure.
FT INIT_MET 0 0
FT ACT_SITE 240 240

FT MUTAGEN 240 240 H->A: LOSS OF ACTIVITY.
SQ SEQUENCE 314 AA; 34194 MW; 21B51B3878A73FA1 CRC64;

Query Match 35.7%; Score 5; DB 1; Length 314;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 AVHDP 6
| | | | |
Db 238 AVHDP 242

Search completed: January 31, 2002, 13:39:20
Job time: 80 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:12 ; Search time 65.13 Seconds
(without alignments)
4.837 Million cell updates/sec

Title: US-08-957-709-75

Perfect score: 14

Sequence: 1 SAVHDPGYEGRPEY 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCRUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	4	US-08-822-774-56
2	5	35.7	24	2	US-08-585-585A-12
3	5	35.7	27	3	US-08-685-808-11
4	5	35.7	27	4	US-08-505-860C-11
5	5	35.7	34	6	5514582-24
6	5	35.7	40	1	US-08-188-228-18
7	5	35.7	40	1	US-08-188-228-24
8	5	35.7	40	1	US-08-332-643-18
9	5	35.7	40	1	US-08-332-643-24
10	5	35.7	40	1	US-08-332-638-18
11	5	35.7	40	1	US-08-332-638-24
12	5	35.7	45	1	US-08-278-089A-25
13	5	35.7	45	2	US-08-838-957A-24
14	5	35.7	78	1	US-08-264-534-2
15	5	35.7	78	1	US-08-083-590A-15
16	5	35.7	78	1	US-08-465-500-2
17	5	35.7	78	2	US-08-346-126-2
18	5	35.7	78	2	US-08-346-128-2
19	5	35.7	78	3	US-08-532-384-15
20	5	35.7	78	3	US-08-893-828-2
21	5	35.7	106	1	US-07-623-611-4
22	5	35.7	106	1	US-07-623-611-7
23	5	35.7	106	5	PCT-US91-09108-4
24	5	35.7	106	5	PCT-US91-09108-7
25	5	35.7	148	1	US-08-253-155A-25
26	5	35.7	154	2	US-08-785-795-4
27	5	35.7	282	4	US-08-750-180-2

28 5 35.7 325 2 US-09-055-097-5 Sequence 5, Appli
29 5 35.7 331 1 US-08-356-180-3 Sequence 3, Appli
30 5 35.7 335 4 US-09-232-191-21 Sequence 21, Appli
31 5 35.7 335 4 US-09-232-200-21 Sequence 21, Appli
32 5 35.7 335 4 US-09-232-197-21 Sequence 21, Appli
33 5 35.7 341 1 US-08-356-180-4 Sequence 4, Appli
34 5 35.7 419 4 US-08-640-906-2 Sequence 17, Appli
35 5 35.7 419 4 US-08-640-906-17 Sequence 2, Appli
36 5 35.7 440 2 US-08-574-959A-2 Sequence 2, Appli
37 5 35.7 440 4 US-09-357-014-2 Sequence 1, Appli
38 5 35.7 448 2 US-08-884-072-1 Sequence 1, Appli
39 5 35.7 448 4 US-09-212-168-1 Sequence 1, Appli
40 5 35.7 512 4 US-09-356-818A-2 Sequence 2, Appli
41 5 35.7 548 1 US-08-247-902A-2 Sequence 2, Appli
42 5 35.7 548 5 PCT-US93-10541-2 Sequence 2, Appli
43 5 35.7 601 1 US-08-458-477A-2 Sequence 2, Appli
44 5 35.7 601 2 US-09-033-153-2 Sequence 2, Appli
45 5 35.7 601 4 US-09-325-430B-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-822-774-56
; Sequence 56, Application US/08822774
; Patent No. 6183997

GENERAL INFORMATION:

APPLICANT: HOGREFE, Holly
TITLE OF INVENTION: Polymerase Enhancing Factor (PEF)
TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
STREET: 1200 G Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,774
FILING DATE: 21-MAR-1997
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: KULIK, David J.
REGISTRATION NUMBER: 36,576
REFERENCE/DOCKET NUMBER: 1486/43163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-822-774-56

Query Match 100.0%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 2.3e-09;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SAVHDPGYEGRPEY 14
|||||


```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: no
; US-08-505-860C-11

Query Match          35.7%; Score 5; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 EGRPE 13
Db 5 EGRPE 9

RESULT 5
5514582-24
; PATENT NO. 5514582
; APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.
; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
; IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 43
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,670
; FILING DATE: 21-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 986,931
; FILING DATE: 08-DEC-1992
; APPLICATION NUMBER: 808,122
; FILING DATE: 16-DEC-1991
; APPLICATION NUMBER: 440,625
; FILING DATE: 22-NOV-1989
; APPLICATION NUMBER: 315,015
; FILING DATE: 23-FEB-1989
; SEQ ID NO: 24:
; LENGTH: 34
5514582-24

Query Match          35.7%; Score 5; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGYEG 10
Db 25 PGYEG 29

RESULT 6
US-08-188-228-18
; Sequence 18, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448

; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-188-228-18

Query Match          35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GYEGR 11
Db 6 GYEGR 10

RESULT 7
US-08-188-228-24
; Sequence 24, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
```

TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-228-24

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GYGR 11
 |||||
DB 6 GYGR 10

RESULT 8
US-08-332-643-18
; Sequence 18, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-332-643-18

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GYGR 11
 |||||
DB 6 GYGR 10

RESULT 9
US-08-332-643-24
; Sequence 24, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-332-643-24

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GYGR 11
 |||||
DB 6 GYGR 10

RESULT 10
US-08-332-638-18
; Sequence 18, Application US/08332638
; Patent No. 5646250
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,638
FILING DATE: 01-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,643
FILING DATE: 17 APR 1992
APPLICATION NUMBER: US/08/049,460
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: No. 5646250and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31340
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-332-638-18

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYGR 11
Db 6 GYGR 10

RESULT 11
US-08-332-638-24
Sequence 24, Application US/08332638
Patent No. 5646250
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,638
FILING DATE: 01-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,643
FILING DATE: 17 APR 1992
APPLICATION NUMBER: US/08/049,460
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: No. 5646250and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31340

TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-332-638-24

Query Match 35.7%; Score 5; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYGR 11
Db 6 GYGR 10

RESULT 12
US-08-278-089A-25
Sequence 25, Application US/08278089A
Patent No. 5681714
GENERAL INFORMATION:
APPLICANT: Breitman, Martin L.
APPLICANT: Rossant, Janet
APPLICANT: Dumont, Daniel J.
TITLE OF INVENTION: No. 5681714el Receptor Tyrosine Kinase
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/278,089A
FILING DATE: 20-JUL-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kurdydk, Linda M.
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 3153-111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
IMMEDIATE SOURCE:
CLONE: No. 5681714ch
US-08-278-089A-25

Query Match 35.7%; Score 5; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYGR 11

Db 37 GYGR 41

RESULT 13

US-08-838-957A-24

Sequence 24, Application US/08838957A

Patent No. 5998187

GENERAL INFORMATION:

APPLICANT: Breitman, Martin L.

APPLICANT: Rossant, Janet

APPLICANT: Dumont, Daniel J.

APPLICANT: Yamaguchi, Terry P.

TITLE OF INVENTION: No. 5998187el Receptor Tyrosine Kinase

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bereskin & Parr

STREET: 40 King Street West

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5H 3Y2

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/838,957A

FILING DATE: 23-APR-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kurodyk, Linda M.

REGISTRATION NUMBER: 34,971

REFERENCE/DOCKET NUMBER: 3153-212

TELEPHONE: (416) 364-7311

TELEFAX: (416) 361-1398

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:

LENGTH: 45 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

IMMEDIATE SOURCE:

CLONE: NO. 5998187ch

US-08-838-957A-24

Query Match 35.7%; Score 5; DB 2; Length 45;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GYGR 11

Db 37 GYGR 41

RESULT 14

US-08-264-534-2

Sequence 2, Application US/08264534

Patent No. 5648464

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon et al.

TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains

TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

US-08-264-534-2

Query Match 35.7%; Score 5; DB 1; Length 78;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PYEG 10

Db 70 PYEG 74

RESULT 15

US-08-083-590A-15

Sequence 15, Application US/08083590A

Patent No. 5786158

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

TITLE OF INVENTION: Nucleic Acids

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/083,590A

FILING DATE: 25-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

US-08-264-534-2

Query Match 35.7%; Score 5; DB 1; Length 78;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PYEG 10

Db 70 PYEG 74

RESULT 15

US-08-083-590A-15

Sequence 15, Application US/08083590A

Patent No. 5786158

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

TITLE OF INVENTION: Nucleic Acids

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/083,590A

FILING DATE: 25-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

US-08-264-534-2


```

; TELEFAX: 212 869864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-083-590A-15

```

```

Query Match      35.7%; Score 5; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      6 PGYEG 10
        |||||
Db      70 PGYEG 74

```

Search completed: January 31, 2002, 13:15:12
Job time: 93 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:18:08 ; Search time 140.03 Seconds
(without alignments)
7.406 Million cell updates/sec

Title: US-08-957-709-75
Perfect score: 14
Sequence: 1 SAVHDPGYEGRPEY 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_GenSeq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	14	100.0	14	AAW72851	Methanococcus jann
2	7	50.0	623	AAW72197	HSV-2 strain SB5 C
3	7	50.0	1196	AAW72105	HSV-2 strain SB5 C
4	6	42.9	418	AAW72105	Partial human IgsF
5	6	42.9	440	AAW72105	A human leukocyte
6	6	42.9	440	AAW72105	Partial human IgsF
7	6	42.9	478	AAW72105	Partial human IgsF
8	6	42.9	613	AAW72105	Human immune syste
9	6	42.9	613	AAW72105	Human EST encoded
10	6	42.9	613	AAW72105	Human secreted pro
11	6	42.9	613	AAW72105	Human secreted pro

12	5	35.7	17	13	AAW25383	Junction sequence
13	5	35.7	24	18	AAW15196	Peptide encoded by
14	5	35.7	27	21	AAW81431	Amino acids encode
15	5	35.7	31	21	AAW26266	pTugAS expression
16	5	35.7	40	18	AAW25649	Rat cadherin-8 cyt
17	5	35.7	40	18	AAW25643	Rat cadherin-11 cy
18	5	35.7	40	18	AAW13114	Rat cadherin-8 cyt
19	5	35.7	40	18	AAW13117	Rat cadherin-11 cy
20	5	35.7	42	20	AAW27029	Amino acid sequenc
21	5	35.7	45	21	AAW59061	Tek receptor Notch
22	5	35.7	63	22	AAW35133	Peptide #9170 enco
23	5	35.7	65	22	AAW47671	Human colon cancer
24	5	35.7	67	22	AAW41970	Human polypeptide
25	5	35.7	78	21	AAW00284	Human secreted pro
26	5	35.7	83	22	AAW00810	C glutamicum prote
27	5	35.7	87	20	AAW31975	HSV-1 ICP47. Herp
28	5	35.7	88	17	AAW92556	Wild-type viral US
29	5	35.7	99	21	AAW01519	Human secreted pro
30	5	35.7	106	13	AAW25148	Bifunctional inhib
31	5	35.7	106	13	AAW25152	Bifunctional inhib
32	5	35.7	113	22	AAW77168	Human colon cancer
33	5	35.7	115	21	AAW54084	Arabidopsis thalia
34	5	35.7	120	21	AAW53841	Human colon cancer
35	5	35.7	148	17	AAW90533	pJG4-5-CDK-BP clon
36	5	35.7	152	21	AAW94877	Human protein clon
37	5	35.7	154	20	AAW28301	Amino acid sequenc
38	5	35.7	155	21	AAW15745	Arabidopsis thalia
39	5	35.7	158	21	AAW54112	Arabidopsis thalia
40	5	35.7	160	22	AAW92667	C glutamicum prote
41	5	35.7	160	22	AAW80182	Corynebacterium gl
42	5	35.7	161	21	AAW43464	Human cancer assoc
43	5	35.7	162	21	AAW25097	Arabidopsis thalia
44	5	35.7	164	21	AAW25096	Arabidopsis thalia
45	5	35.7	164	21	AAW47068	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAW72851
ID AAW72851 standard; Peptide; 14 AA.
XX
AC AAW72851;
XX
XX
DT 01-MAR-1999 (first entry)
XX
XX Methanococcus jannaschii dCTP deaminase uridine-binding motif.
DE
XX
XX Polymerase enhancing factor; PEF; dCTP deaminase; PCR;
KW amplification; sequencing; replication.
XX
XX Methanococcus jannaschii.
OS
XX
XX WO9842860-A1.
PN
XX
XX 01-OCT-1998.
PD
XX
XX 20-MAR-1998; 98WO-US05497.
PF
XX
XX 24-OCT-1997; 97US-0957709.
PR
XX
XX 21-MAR-1997; 97US-0822774.
PR
XX
XX (STRA-) STRATAGENE.
PA
XX
XX Hansen CJ, Hogrefe H;
PI
XX
XX WPI; 1998-542284/46.
DR
XX
XX Polymerase enhancing factor proteins, extracts and complexes -
PT improve the polymerisation activity of nucleic acid polymerases, for
PT use in amplification, sequencing and replication
XX

PS Claim 71; Page 47; 161pp; English.

XX This is the uridine-binding motif of the dCTP deaminase of

CC Methanococcus jannaschii. Sequences are provided (see AAW72849-57)

CC of the uridine-binding motifs of dUTPases and dCTP deaminases of

CC Pyrococcus furiosus (see AAW72847), Methanococcus jannaschii,

CC Desulfurolobus ambivalens, Escherichia coli, yeast, human and

CC herpesvirus; a consensus (see AAW72848) is also provided. A claimed

CC method of enhancing a nucleic acid polymerase reaction comprises

CC performing the reaction in the presence of one or more of the

CC following: a polymerase enhancing factor (PEF), a dUTPase, a

CC protein that turns-over dUTP and a protein having one or more of

CC the sequences provided in AAW72848-57. A claimed protein having PEF

CC activity comprises one or more of sequences given in AAW72848-57.

CC kits are provided for replicating nucleic acids, for site-directed

CC mutagenesis, for nucleic acid sequencing or for amplification

CC (preferably PCR or RT-PCR).

XX Sequence 14 AA:

SAVHPDGYEGRPEY 14
 |||||

SAVHPDGYEGRPEY 14
 |||||

RESULT 2

AAW72197

ID AAW72197 standard; Protein: 623 AA.

XX

AC AAW72197;

XX

DT 13-JAN-1999 (first entry)

XX

DE HSV-2 strain SB5 Contig ID 15 ORF#32 protein.

XX

DE HSV-2 strain SB5; immunological response induction; therapy;

KW antiviral identification; viral protein inhibitor.

KW

KW Herpes simplex virus type 2.

OS

OS WO9820016-A1.

PN

PN 14-MAY-1998.

PD

PD 31-OCT-1997; 97WO-US20016.

PF

PF 09-JUN-1997; 97US-0049018.

XX

XX 04-NOV-1996; 96US-0030279.

PR

PR (SMIK) SMITHKLINE BEECHAM CORP.

XX

XX Chan JY, Dabrowski-Anaral CE, Delvecchio AM, Dillon SB;

PI Esser KM, Leary JJ;

PI

XX WPI: 1998-286847/25.

DR

DR N-PSDB; AAV62176.

XX

XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention

PT and treatment of infection or inducing immunological response in

PT mammal

XX

XX Claim 10; Page 123; 748pp; English.

XX

XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein

CC sequence of the invention. This sequence was isolated from a HSV-2 strain

CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.

CC Based on homology, this sequence is a major DNA-binding protein.

CC The proteins can be used for the treatment or prevention of disease, to

CC induce an immunological response in a mammal or to identify inhibitors,

CC activators or novel antivirals. Antagonists of the proteins can be used

CC to inhibit a viral polypeptide. The DNA sequence or a vector containing

CC it can also be used to induce an immunological response in a mammal.

XX Sequence 623 AA:

Query Match 50.0%; Score 7; DB 19; Length 623;

Best Local Similarity 100.0%; Pred. No. 2.6;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGYEGRP 12
 |||||

Db 428 pgyegrp 434
 |||||

RESULT 3

AAW72105

ID AAW72105 standard; Protein: 1196 AA.

XX

AC AAW72105;

XX

DT 18-DEC-1998 (first entry)

XX

DE HSV-2 strain SB5 Contig ID 12 ORF#2 protein.

XX

XX HSV-2 strain SB5; immunological response induction; therapy;

KW antiviral identification; viral protein inhibitor.

KW

KW Herpes simplex virus type 2.

OS

OS WO9820016-A1.

PN

PN 14-MAY-1998.

PD

PD 31-OCT-1997; 97WO-US20016.

PF

PF 09-JUN-1997; 97US-0049018.

XX

XX 04-NOV-1996; 96US-0030279.

PR

PR (SMIK) SMITHKLINE BEECHAM CORP.

XX

XX Chan JY, Dabrowski-Anaral CE, Delvecchio AM, Dillon SB;

PI Esser KM, Leary JJ;

PI

XX WPI: 1998-286847/25.

DR

DR N-PSDB; AAV62156.

XX

XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention

PT and treatment of infection or inducing immunological response in

PT mammal

XX

XX Claim 10; Page 83-84; 748pp; English.

XX

XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein

CC sequence of the invention. This sequence was isolated from a HSV-2 strain

CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 12.

CC The proteins can be used for the treatment or prevention of disease, to

CC induce an immunological response in a mammal or to identify inhibitors,

CC activators or novel antivirals. Antagonists of the proteins can be used

CC to inhibit a viral polypeptide. The DNA sequence or a vector containing

CC it can also be used to induce an immunological response in a mammal.

XX Sequence 1196 AA:

Query Match 50.0%; Score 7; DB 19; Length 1196;

Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGYEGRP 12
 |||||

CC with Alzheimer's disease and Huntington's chorea, and inhibit growth and
 XX infection of agents such as bacteria, fungi, viruses and other parasites.
 SQ Sequence 418 AA;

Query Match 42.9%; Score 6; DB 22; Length 418;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGVEGR 11
 Db 177 pygegr 182

RESULT 5
 AAB08766
 ID AAB08766 standard; Protein; 440 AA.
 XX
 AC AAB08766;
 XX
 XX 02-JAN-2001 (first entry)
 XX
 DE A human leukocyte and blood related protein (LBAP).
 XX
 KW Human; leukocyte and blood related protein; LBAP; arteriosclerosis;
 KW cell proliferative disorder; actinic keratosis; atherosclerosis;
 KW bursitis; cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
 KW myelofibrosis; paroxysmal nocturnal hemoglobinuria; cancer;
 KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
 KW teratocarcinoma; autoimmune disorder; inflammatory disorder;
 KW acquired immunodeficiency syndrome; AIDS; Addison's disease;
 KW adult respiratory distress syndrome; allergy; ankylosing spondylitis;
 KW amyloidosis; anaemia; asthma; autoimmune haemolytic anaemia; infection;
 KW Werner syndrome; haemodialysis; extracorporeal circulation; trauma.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Modified-site 2
 FT /note= "potential phosphorylation site"
 FT Domain 6..99
 FT /note= "immunoglobulin domain"
 FT Modified-site 19
 FT /note= "potential phosphorylation site"
 FT Modified-site 32
 FT /note= "potential phosphorylation site"
 FT Modified-site 42
 FT /note= "potential phosphorylation site"
 FT Modified-site 77
 FT /note= "potential phosphorylation site"
 FT Modified-site 98
 FT /note= "potential phosphorylation site"
 FT Domain 146..235
 FT /note= "immunoglobulin domain"
 FT Modified-site 154
 FT /note= "potential glycosylation site"
 FT Modified-site 215
 FT /note= "potential phosphorylation site"
 FT Modified-site 230
 FT /note= "potential phosphorylation site"
 FT Modified-site 231
 FT /note= "potential phosphorylation site"
 FT Modified-site 251
 FT /note= "potential phosphorylation site"
 FT Domain 282..373
 FT /note= "immunoglobulin domain"
 FT Modified-site 290
 FT /note= "potential glycosylation site"
 FT Modified-site 292
 FT /note= "potential phosphorylation site"
 FT Modified-site 353
 FT /note= "potential phosphorylation site"

CC 429 pygegrp 435
 XX
 SQ Sequence 418 AA;

Query Match 42.9%; Score 6; DB 22; Length 418;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGVEGR 11
 Db 177 pygegr 182

RESULT 4
 AAB81410
 ID AAB81410 standard; Protein; 418 AA.
 XX
 AC AAB81410;
 XX
 XX 18-JUN-2001 (first entry)
 XX
 DE Partial human IgSF protein, SEQ ID NO: 3.
 XX
 KW Human; IgSF; Immunoglobulin superfamily; immunostimulant;
 KW immunosuppressive; cytostatic; antirheumatic; antiarthritic; nootropic;
 KW antidiabetic; osteopathic; vulnery; antiinflammatory; neuroprotective;
 KW anticonvulsant; antimicrobial; gene therapy; immune disorder;
 KW autoimmune disorder; infection; haematopoiesis; osteoporosis;
 KW periodontal disease; inflammation; nervous system disorder.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Domain 1..77
 FT /label= Domain_1
 FT Region 1..75
 FT /note= "cysteine-bonded loop region"
 FT Domain 124..213
 FT /label= Domain_2
 FT Disulfide-bond 131..211
 FT /note= "forms loop region"
 FT Domain 260..351
 FT /label= Domain_3
 FT Disulfide-bond 267..349
 FT /note= "forms loop region"
 XX
 PN WO200127278-A2.
 XX
 XX 19-APR-2001.
 XX
 XX 16-OCT-2000; 2000WO-US28661.
 XX
 XX 14-OCT-1999; 99US-0417791.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Ford J, Yeung G;
 XX
 XX WPI; 2001-273774/28.
 XX
 XX N-PSDB; AAF82525.
 XX
 XX Polynucleotide sequence encodes a novel member of the Immunoglobulin
 XX superfamily which can be used in treatment of immune deficiencies and
 XX disorders, promotion of wound healing and prevention of microbial
 XX infections -
 XX
 XX Claim 8; Page 110-112; 120pp; English.
 XX
 CC The present sequence is a novel member of the Immunoglobulin (Ig)
 CC superfamily (IgSF). IgSF polypeptides and polynucleotides encoding
 CC them can be used for the prevention and treatment of disorders involving
 CC aberrant protein expression or biological activity. Members of the
 CC immunoglobulin superfamily are involved in cell attachment, cell-cell
 CC recognition, cell growth and replication, cell shape, cell migration,
 CC immunological responses, molecular transport, shape recognition, toxin
 CC neutralisation, regulation of gene transcription and cell death marking.
 CC The IgSF polypeptides can be used for the treatment of immune
 CC deficiencies, autoimmune disorders and disorders caused by viral or
 CC fungal infections. Such disorders include multiple sclerosis, rheumatoid
 CC arthritis, cancer and insulin dependent diabetes mellitus. They can also
 CC be used to regulate haematopoiesis, treat osteoporosis and periodontal
 CC diseases, promote wound healing, treat inflammatory disorders such as
 CC arthritis, treat nervous system disorders including lesions associated

```

FT Modified-site 389
FT Domain /note= "potential phosphorylation site"
FT 410..434
FT /note= "transmembrane domain"
XX
XX WO2000052161-A2.
XX
XX 08-SEP-2000.
XX
XX 29-FEB-2000; 2000WO-US05153.
XX
XX 01-MAR-1999; 99US-0122080.
XX (INCYTE) INCYTE PHARM INC.
XX
XX Lal P, Yue H, Hillman JL, Lu DAM, Baughn MR, Tang YT, Azimzai Y;
XX WPI: 2000-587310/55.
XX N-PSDB; AAA64686.
XX
XX Leukocyte and blood associated proteins and polynucleotides encoding
XX them, useful for diagnosis, treatment and prevention of
XX autoimmune/inflammatory disorders and cell proliferative disorders
XX including cancer -
XX
XX Claim 1; Page 66-67; 70pp; English.
XX
XX The present sequence presents a human leukocyte and blood related
XX protein, designated LBAP. LBAP polynucleotides and polypeptides are
XX useful for treating or preventing a disorder associated with decreased
XX expression or activity of LBAP including a cell proliferative disorder
XX such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis,
XX cirrhosis, hepatitis, mixed connective tissue disease (MCTD),
XX myelofibrosis, paroxysmal nocturnal hemoglobinuria, etc., cancers
XX including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma,
XX sarcoma, teratocarcinoma and in particular cancers of the adrenal
XX gland, bladder, bone, bone marrow, brain, breast, cervix, etc., and
XX an autoimmune/inflammatory disorder such as acquired immunodeficiency
XX syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,
XX allergies, ankylosing spondylitis, amyloidosis, anaemia, asthma,
XX atherosclerosis, autoimmune haemolytic anaemia, etc., Werner syndrome,
XX complications of cancer, haemodialysis, and extracorporeal circulation,
XX viral, bacterial, fungal, parasitic, protozoan, and helminthic
XX infections, and trauma.
XX
XX Sequence 440 AA;

Query Match 42.9%; Score 6; DB 21; Length 440;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGYEGR 11
Db 199 pgyegr 204
|||||

RESULT 6
AAB81412
ID AAB81412 standard; Protein; 440 AA.
XX
XX AAB81412;
XX
XX 18-JUN-2001 (first entry)
XX
XX Partial human IgSF protein, SEQ ID NO: 7.
XX
XX Human; IgSF; immunoglobulin superfamily; immunostimulant;
XX immunosuppressive; cytostatic; antirheumatic; antiarthritic; neutropic;
XX antidiabetic; osteopathic; vulnary; antinflammatory; neuroprotective;
XX anticonvulsant; antimicrobial; gene therapy; immune disorder;
XX autoimmune disorder; infection; haematopoiesis; osteoporosis;
XX periodontal disease; inflammation; nervous system disorder.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 1..99
XX Region 1..97
XX Domain /note= "cysteine-bonded loop region"
XX 146..235
XX /label= Domain_2
XX Disulfide-bond 153..233
XX /note= "forms loop region"
XX Domain 282..373
XX /label= Domain_3
XX Disulfide-bond 289..371
XX /note= "forms loop region"
XX
XX WO200127278-A2.
XX
XX 19-APR-2001.
XX
XX 16-OCT-2000; 2000WO-US28661.
XX
XX 14-OCT-1999; 99US-0417791.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Ford J, Yeung G;
XX WPI: 2001-273774/28.
XX N-PSDB; AAF82527.
XX
XX Polynucleotide sequence encodes a novel member of the Immunoglobulin
XX superfamily which can be used in treatment of immune deficiencies and
XX disorders, promotion of wound healing and prevention of microbial
XX infections -
XX
XX Claim 8; Page 118-119; 120pp; English.
XX
XX The present sequence is a novel member of the Immunoglobulin (Ig)
XX superfamily (IGSF). IGSF polypeptides and polynucleotides encoding
XX them can be used for the prevention and treatment of disorders involving
XX aberrant protein expression or biological activity. Members of the
XX immunoglobulin superfamily are involved in cell attachment, cell-cell
XX recognition, cell growth and replication, cell shape, cell migration,
XX immunological responses, molecular transport, shape recognition, toxin
XX neutralisation, regulation of gene transcription and cell death marking.
XX The IGSF polypeptides can be used for the treatment of immune
XX deficiencies, autoimmune disorders and disorders caused by viral or
XX fungal infections. Such disorders include multiple sclerosis, rheumatoid
XX arthritis, cancer and insulin dependent diabetes mellitus. They can also
XX be used to regulate haematopoiesis, treat osteoporosis and periodontal
XX diseases, promote wound healing, treat inflammatory disorders such as
XX arthritis, treat nervous system disorders including lesions associated
XX with Alzheimer's disease and Huntington's chorea, and inhibit growth and
XX infection of agents such as bacteria, fungi, viruses and other parasites.
XX
XX Sequence 440 AA;

Query Match 42.9%; Score 6; DB 22; Length 440;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGYEGR 11
Db 199 pgyegr 204
|||||

RESULT 7
AAB81411
ID AAB81411 standard; Protein; 478 AA.
XX

```

AAB81411;
 18-JUN-2001 (first entry)
 Partial human IgSF protein, SEQ ID NO: 5.
 Human; IgSF; immunoglobulin superfamily; immunostimulant;
 immunosuppressive; cytostatic; antirheumatic; antiarthritic; nootropic;
 antidiabetic; osteopathic; vulnerary; antiinflammatory; neuroprotective;
 anticonvulsant; antimicrobial; gene therapy; immune disorder;
 autoimmune disorder; infection; haematopoiesis; osteoporosis;
 periodontal disease; inflammation; nervous system disorder.
 Homo sapiens.
 Key Location/Qualifiers
 Domain 1..137 /label= Domain_1
 Region 1..135 /note= "cysteine-bonded loop region"
 Domain 184..273 /label= Domain_2
 Disulfide-bond 191..271 /note= "forms loop region"
 Domain 320..411 /label= Domain_3
 Disulfide-bond 327..409 /note= "forms loop region"
 WO200127278-A2.
 19-APR-2001.
 16-OCT-2000; 2000WO-US28661.
 14-OCT-1999; 99US-0417791.
 (HYSE-) HYSEQ INC.
 Ford J, Yeung G;
 WPI: 2001-273774/28.
 N-PSDB; AAF82526.
 Polynucleotide sequence encodes a novel member of the immunoglobulin
 superfamily which can be used in treatment of immune deficiencies and
 disorders, promotion of wound healing and prevention of microbial
 infections -
 Claim 8; Page 114-115; 120pp; English.
 The present sequence is a novel member of the immunoglobulin (Ig)
 superfamily (IgSF). IgSF polypeptides and polynucleotides encoding
 them can be used for the prevention and treatment of disorders involving
 aberrant protein expression or biological activity. Members of the
 immunoglobulin superfamily are involved in cell attachment, cell-cell
 recognition, cell growth and replication, cell shape, cell migration,
 immunological responses, molecular transport, shape recognition, toxin
 neutralisation, regulation of gene transcription and cell death marking.
 The IgSF polypeptides can be used for the treatment of immune
 deficiencies, autoimmune disorders and disorders caused by viral or
 fungal infections. Such disorders include multiple sclerosis, rheumatoid
 arthritis, cancer and insulin dependent diabetes mellitus. They can also
 be used to regulate haematopoiesis, treat osteoporosis and periodontal
 diseases, promote wound healing, treat inflammatory disorders such as
 arthritis, treat nervous system disorders including lesions associated
 with Alzheimer's disease and Huntington's chorea, and inhibit growth and
 infection of agents such as bacteria, fungi, viruses and other parasites.
 Sequence 478 AA;

Query Match 42.9%; Score 6; DB 22; Length 478;

Best Local Similarity 100.0%; Pred. No. 28;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 PGYEGR 11
 Db 237 PGYEGR 242
 RESULT 8
 AAB15536
 ID AAB15536 standard; Protein; 613 AA.
 XX
 AC AAB15536;
 XX
 DT 28-FEB-2001 (first entry)
 XX
 DE Human immune system molecule from Incyte clone 2705028.
 XX
 KW Anti-inflammatory; keratolytic; anti-HIV; anti-allergic; antianaemic;
 antiarteriosclerotic; antiasthmatic; antidiabetic; nephrotropic; cancer;
 antigout; dermatological; antithyroid; virucide; hepatotropic; antibody;
 immunosuppressive; cytostatic; fungicide; protozoacide; antibacterial;
 gene therapy; diagnostic; immunological disorder; viral infection;
 bacterial infection; fungal infection; parasitic infection; immunogen.
 Homo sapiens.
 OS
 XX WO200060080-A2.
 PN
 XX 12-OCT-2000.
 XX
 XX 04-APR-2000; 2000WO-US09072.
 XX
 XX 05-APR-1999; 99US-0127852.
 PR
 XX 05-MAY-1999; 99US-0132647.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Yue H, Lal P, Tang YT, Baughn MR, Azimzai Y, Lu DAM;
 PI
 XX WPI: 2000-665005/64.
 DR
 DR N-PSDB; AAA95775.
 XX
 XX New human immune system molecules 1-15 and polynucleotides encoding
 them useful for diagnosing, treating or preventing e.g. immunological
 disorders, infections, cell proliferative disorders, microbial
 infections -
 Claim 1; Page 75-76; 95pp; English.
 This sequence represents a human immune system molecule (IMOL) encoded
 by the cDNA isolated as clone 2705028 from the Incyte PONSAT01 library.
 The human IMOLs (AAB1536-B15550) and their encoding polynucleotides
 (AAA95775-A95789), and compositions comprising them are useful for the
 diagnosis, treatment or prevention of immunological disorders,
 infections and cell proliferative disorders, including cancer. The IMOL
 may be used to treat or prevent disorders associated with decreased
 expression or activity of IMOL, such as immunological disorders
 (e.g. inflammation, actinic keratosis, AIDS, Addison's disease),
 haematopoietic cancer, infections caused by virus (e.g. adenovirus,
 parvovirus, coronavirus), bacteria (e.g. Staphylococcus, Streptococcus,
 Shigella), fungi (e.g. Aspergillus, Blastomyces), parasites (e.g.
 Plasmodium, Trypanosoma, intestinal protozoa), cell proliferative
 disorders (e.g. actinic keratosis, arteriosclerosis, bursitis), and
 cancers (e.g. leukemia, melanoma, sarcoma). The peptides are also
 useful as immunogens for the development of antibodies that
 specifically recognize these peptides. The polynucleotides may be used
 to detect and quantify gene expression in biopsied tissues in which
 expression of IMOL may be correlated with the disease, as targets in a
 microarray, to detect differences in gene sequences among normal,
 carrier and affected individuals, and for screening libraries of
 compounds in drug screening techniques. Antibodies which specifically
 bind to IMOL may be used for the diagnosis of disorders characterized

CC by expression of IMOL, or in assays to monitor patients being treated
 CC with IMOL or agonists, antagonists, or inhibitors of IMOL.

SQ Sequence 613 AA;

Query Match 42.9%; Score 6; DB 21; Length 613;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGYEGR 11
 |||||
 Db 372 pgyegr 377

RESULT 9
 AAM24248
 ID AAM24248 standard; Protein: 613 AA.

XX AC AAM24248;

XX DT 12-OCT-2001 (first entry)

XX DE Human EST encoded protein SEQ ID NO: 1773.

XX KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition.

XX OS Homo sapiens.

XX PN WO200154477-A2.

XX PD 02-AUG-2001.

XX PF 25-JAN-2001; 2001WO-US02687.

XX PR 25-JAN-2000; 2000US-0491404.

XX PR 17-JUL-2000; 2000US-0617746.

XX PR 03-AUG-2000; 2000US-0631451.

XX PR 15-SEP-2000; 2000US-0663870.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI; 2001-476164/51.

XX N-PSDB; AAH98907.

XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 antibodies and research use -

XX PS Claim 20; Page 1164-1165; 1275pp; English.

XX CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.

XX SQ Sequence 613 AA;

Query Match 42.9%; Score 6; DB 22; Length 613;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGYEGR 11

Db 372 pgyegr 377
 |||||

RESULT 10

AAB90544

ID AAB90544 standard; Protein: 613 AA.

XX AC AAB90544;

XX DT 01-JUN-2001 (first entry)

XX DE Human secreted protein, SEQ ID NO: 82.

XX KW Human; secreted protein; immunomodulatory; antisclerotic;
 KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
 KW neotropic; anticonvulsant; antialzheimers; antiparkinsonian;
 KW antimicrobial; vulnerrary; vaccine; gene therapy; cancer;
 KW protein coordinate data; infection.

XX OS Homo sapiens.

XX PN WO200121658-A1.

XX PD 29-MAR-2001.

XX PF 22-SEP-2000; 2000WO-US26013.

XX PR 24-SEP-1999; 99US-0155709.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;

XX PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;

XX PI Young PE, Wei P, Florence KA;

XX WPI; 2001-235311/24.

XX N-PSDB; AAF97884.

XX PT Nucleic acids encoding 32 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -

XX PS Claim 11; Page 763-764; 890pp; English.

XX CC The present sequence is one of 32 novel human secreted polypeptides. The
 CC nucleic acid molecules and polypeptides may be used in the prevention,
 CC diagnosis and treatment of diseases such as immune disorders (e.g.
 CC multiple sclerosis, systemic lupus erythematosus and human
 CC immuno-deficiency virus (HIV) infections), hyperproliferative disorders
 CC (e.g. cancers and Gaucher's disease), cardiovascular diseases
 CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC and/or chemotaxis. The nucleic acid molecules may be used to produce the
 CC secreted polypeptides. They may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples. The polypeptides may be used as antigens in the
 CC production of antibodies and in assays to identify modulators of
 CC their expression and activity.

XX SQ Sequence 613 AA;

Query Match 42.9%; Score 6; DB 22; Length 613;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGYEGR 11
 |||||

Db 372 pgyegr 377

RESULT 11

AAB90560
ID AAB90560 standard; Protein; 613 AA.

XX AC AAB90560;

XX DT 01-JUN-2001 (first entry)

XX DE Human secreted protein, SEQ ID NO: 98.

XX KW Human; secreted protein; immunomodulatory; antisclerotic;
XX KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
XX KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
XX KW neurotropic; anticonvulsant; antiparkinsonian;
XX KW antimicrobial; vulnerrary; vaccine; gene therapy; cancer;
XX KW protein coordinate data; infection.

XX OS Homo sapiens.

XX PN WO200121658-A1.

XX PD 29-MAR-2001.

XX PF 22-SEP-2000; 2000WO-US26013.

XX PR 24-SEP-1999; 99US-0155709.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
XX PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
XX PI Young PE, Wei P, Florence KA;
XX DR WPI; 2001-235311/24.
XX DR N-PSDB; AAF97900.

XX PT Nucleic acids encoding 32 human secreted polypeptides, useful for
XX PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX PT disease and diabetic retinopathy -

XX PS Claim 11; Page 784-786; 890pp; English.

XX CC The present sequence is one of 32 novel human secreted polypeptides. The
XX CC nucleic acid molecules and polypeptides may be used in the prevention,
XX CC diagnosis and treatment of diseases such as immune disorders (e.g.
XX CC multiple sclerosis, systemic lupus erythematosus and human
XX CC immunodeficiency virus (HIV) infections), hyperproliferative disorders
XX CC (e.g. cancers and Gaucher's disease), cardiovascular diseases
XX CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
XX CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
XX CC neovascularisation and diabetic retinopathy), neurological disorders
XX CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
XX CC infectious diseases and/or for promoting wound healing, regeneration
XX CC and/or chemotaxis. The nucleic acid molecules may be used to produce the
XX CC secreted polypeptides. They may also be used as DNA probes in diagnostic
XX CC assays to detect and quantitate the presence of similar nucleic acid
XX CC sequences in samples. The polypeptides may be used as antigens in the
XX CC production of antibodies and in assays to identify modulators of
XX CC their expression and activity.

XX SQ Sequence 613 AA;

Query Match 42.9%; Score 6; DB 22; Length 613;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 PGYEGR 11

Db 372 pgyegr 377

RESULT 12

AAR25383
ID AAR25383 standard; Protein; 17 AA.

XX AC AAR25383;

XX DT 04-JAN-1993 (first entry)

XX DE Junction sequence of male-applig fusion protein.

XX KW bifunctional inhibitor; platelet activation; thrombin; inhibit clot;
XX KW formation; accretion; fibrin deposition; myocardial infarction;
XX KW thrombosis; increasing reocclusion time; decreasing reperfusion;
XX KW time; inhibiting metastatic cell growth.

XX OS Synthetic.

XX FH Key Location/Qualifiers
XX FT Cleavage-site 7..8
XX FT /label= Factor Xa

XX PN WO9210575-A.

XX PD 25-JUN-1992.

XX PF 05-DEC-1991; 91WO-US09108.

XX PR 07-DEC-1990; 90US-0623611.

XX PA (BIOJ) BIOGEN INC.

XX PI Chao BH, Maraganore JM, Strauch KL, Thompson JS;

XX DR WPI; 1992-234630/28.

XX DR N-PSDB; AAQ25890.

XX PT Bi: functional inhibitors of thrombin and platelet activation -
XX PT comprise glycoprotein IIB or IIIa and thrombin inhibitory
XX PT moieties, for treating thrombotic diseases, atherosclerosis,
XX PT cancer and neurodegenerative conditions

XX PS Example 4; Fig 10; 103pp; English.

XX CC This sequence represents the junction between the male and applig
XX CC portions of the male-applig fusion protein - produced as described
XX CC in AAQ25890.

XX CC See also AAR25142-54 AAR25382,3, AAQ25314,5 AAQ25884-90, AAQ31179,80.

XX SQ Sequence 17 AA;

Query Match 35.7%; Score 5; DB 13; Length 17;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 EGRPE 13

Db 11 egrpe 15

RESULT 13

AAW15196
ID AAW15196 standard; Peptide; 24 AA.

XX AC AAW15196;

XX DT 01-NOV-1997 (first entry)

XX DE Peptide encoded by vector pTugAS.

XX KW Oligosaccharide binding domain; cellulose binding domain;

KW cellulase; endoglucanase C; Cenc; purification; phase partition;
 KW two-phase system; affinity phase separation;
 KW extractive bioconversion; vector; ptugAS.

OS Synthetic.

PH Key Location/Qualifiers
 FT Cleavage-site 2..3
 FT /label- FXa

XX WO9713841-A1.

XX 17-APR-1997.

XX 24-JUL-1996; 96WO-US12282.

XX 24-JUL-1995; 95US-0505860.

XX (UYBR-) UNIV BRITISH COLUMBIA.

XX Haynes CA, Kilburn DG, Tomme P;

XX WPI; 1997-235878/21.

XX N-PSDB; AAT65088.

XX Purifying a polypeptide containing a polysaccharide binding peptide
 PT - using a phase-forming oligosaccharide and separating the phases,
 PT useful for large scale purification of proteins and enrichment of
 PT cell types

XX Disclosure; Fig 3A; 87pp; English.

XX Two peptides (AAW15195 and AAW15196) are respectively encoded by a
 CC nucleotide sequence located upstream of the SacI site (see AAW65087),
 CC and by the SacI-HindIII region (see AAW65088); of vector ptugAS.
 CC This vector can be used for expression of cellulose-binding domain
 CC N1 (AAW15193) of Cellulomonas fimi endoglucanase C (Cenc) in
 CC Escherichia coli. N1 can be produced for use in claimed methods of
 CC phase separation for large-scale protein purification and cell
 CC enrichment.

XX Sequence 24 AA;

Query Match 35.7%; Score 5; DB 18; Length 24;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 EGRPE 13

Db 2 egrpe 6

RESULT 14

AAW81431

ID AAW81431 standard; peptide; 27 AA.

XX AAW81431;

XX 19-JUN-2000 (first entry)

XX Amino acids encoded by 3' portion of ptugAS vector SacI-HindIII region.

XX ptugAS vector; Cellulomonas fimi; endoglucanase C; CBD;
 KW cellulose binding domain; polysaccharide binding peptide;
 KW fusion protein; recombinant protein; cell separation;
 KW affinity phase separation.

XX Synthetic.

XX US6048715-A.

XX 11-APR-2000.

PD

XX 24-JUL-1996; 96US-0685808.

XX 08-JUL-1988; 88US-0216794.

XX 08-APR-1992; 92US-0865095.

XX 25-OCT-1990; 90US-0603967.

XX 24-MAY-1994; 94US-0249037.

XX 24-JUL-1995; 95US-0505860.

XX (UYBR-) UNIV BRITISH COLUMBIA.

XX Haynes CA, Tomme P, Kilburn DG;

XX WPI; 2000-328038/28.

XX Two-phase partition affinity separation system useful for separating
 PT and purifying proteins comprises a phase-forming oligosaccharide
 PT polymer and a phase-separation agent

XX Disclosure; Fig 3A; 46pp; English.

XX The invention relates to a novel two-phase partition system for affinity
 CC separation which comprises a phase-forming oligosaccharide polymer to
 CC which a polysaccharide binding peptide (PBP) binds, and a phase
 CC separation inducing agent. The system may be used for the separation and
 CC purification of recombinant proteins from cell cultures. The protein of
 CC interest is expressed in the host cell as a fusion protein with a
 CC polysaccharide binding peptide, and the cell lysate or fermentation
 CC broth (or other composition comprising the fusion protein) is contacted
 CC with the phase separation system. The fusion protein partitions into the
 CC oligosaccharide polymer phase, and can then be isolated e.g., via the
 CC use of a solution with a low ionic strength, high pH or containing a
 CC chaotropic agent, or by the use of chemical cleavage agents such as
 CC cyanogen bromide. If the fusion protein has been engineered such that
 CC there is a protease recognition site between the PBP and the protein of
 CC interest, the protein can be cleaved from the PBP, which remains bound
 CC to the oligosaccharide. Proteins that may be purified using the method
 CC of the invention include a wide variety of medically and industrially
 CC important proteins e.g., interleukin 2, factor X, insulin, ligninase.
 CC The system can also be used for cell separation and/or enrichment of a
 CC particular cell type; for example, a fusion protein comprising a PBP and
 CC a ligand of a particular receptor can be used to isolate cells
 CC comprising that receptor. The system is useful for concentrating a
 CC component in a mixture, removing contaminants and for preparing solid
 CC state reagents for diagnostic assays. Oligosaccharide polymers can be
 CC obtained inexpensively and the water-soluble cellulosic substrates offer
 CC a new, cost-effective, highly flexible affinity partition system for
 CC continuous purification. Selective binding of PBP from the
 CC oligosaccharide polymer makes it suitable for purification of a wide
 CC variety of compounds using a single oligosaccharide polymer phase
 CC separation system. Hence it is unnecessary to prepare separate systems
 CC for each compound to be separated. Sequences AAW81430-Y81431 represent
 CC the amino acids sequences encoded by the 5' and 3' portions of the
 CC SacI-HindIII region of the ptugAS vector. DNA encoding the Cellulomonas
 CC fimi endoglucanase C cellulose binding domain (CBD) N1 (AAW81423) was
 CC inserted between the 5' and 3' portions of the SacI-HindIII region of the
 CC vector.

XX Sequence 27 AA;

Query Match 35.7%; Score 5; DB 21; Length 27;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 EGRPE 13

Db 5 egrpe 9

RESULT 15

AAB26266

ID AAB26266 standard; Protein; 31 AA.

XX AAB26266;
AC
XX
DT 10-JAN-2001 (first entry)
XX
DE pTugAS expression vector peptide fragment.
XX
KW SLF-CBD; murine; steel factor; pTugAS; cellulose binding domain;
KW exoglucanase; Cex; cell division; differentiation; growth factor;
KW wound healing.
XX
OS Unidentified.
XX
PN WO2000050570-A2.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-CA00162.
XX
PR 23-FEB-1999; 99US-0256499.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Kilburn DG, Jervis E, Doheny JG, Haynes CA;
XX
DR WPI; 2000-611306/58.
DR N-PSDB; AAA93095.
XX
XX Composition for enhancing rate of division and differentiation of
PT growth factor dependent cells, comprising a growth factor receptor
PT complex and its immobilized ligand -
XX
PS Disclosure; Fig 3A; 139pp; English.
XX
CC The fusion protein SLF-CBD comprises the extracellular domain of murine
CC steel factor (SLF) linked to the cellulose binding domain (CBD) of the
CC Cellulomonas fimi exoglucanase Cex. The present sequence is a peptide
CC fragment of pTugAS expression vector. pTugAS can be used for high level
CC expression of SLF-CBD fusion proteins in Escherichia coli. The present
CC sequence comprises the SacI to HindIII region. The present invention
CC relates to the use of SLF-CBD as a component of a composition used for
CC enhancing the rate of division and differentiation of growth factor
CC dependent cells. The composition of the present invention also comprises
CC of a complex of viable cells, containing biologically active growth
CC factor receptors and ligands of the receptors which are immobilised on a
CC biocompatible solid phase via a polysaccharide binding domain i.e.
CC SLF-CBD, (which is diffused on the solid phase), so that cell surface
CC localised growth factor receptor diffuses within the cell surface. The
CC composition of the present invention may be used to enhance wound healing
CC via a growth factor-CBD conjugate, which would stimulate local
CC proliferation of cells involved in wound healing, and/or is chemotactic
CC for such cells.
XX
SQ Sequence 31 AA;

Query Match 35.7%; Score 5; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 EGRPE 13
Db 9 egrpe 13
|||||

Search completed: January 31, 2002, 13:18:10
Job time: 171 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:37:43 ; Search time 130.99 Seconds
(without alignments)
15.633 Million cell updates/sec

Title: US-08-957-709-76
Perfect score: 14
Sequence: 1 PTIVDAGFEGOLTI 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTEMBL17.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-virus.*
13: sp-invertebrate.*
14: sp-unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	158	12	071028 archaeal vi
2	7	50.0	89	2	Q9RB12
3	7	50.0	328	9	Q9G084
4	7	50.0	488	10	Q9LK94
5	6	42.9	110	2	P74723
6	6	42.9	118	2	Q49668
7	6	42.9	184	9	O48474
8	6	42.9	202	5	Q9GY57
9	6	42.9	218	5	Q22723
10	6	42.9	218	11	Q9P8X3
11	6	42.9	260	2	Q9ABU6
12	6	42.9	282	2	O52068
13	6	42.9	290	4	Q9HAQ5
14	6	42.9	325	1	Q9HP97
15	6	42.9	349	2	Q9RW33
16	6	42.9	386	3	Q9V726
17	6	42.9	398	2	Q9AQH6
18	6	42.9	406	2	Q9KL05
19	6	42.9	412	2	O69367

20	6	42.9	430	5	Q9VW98
21	6	42.9	437	3	Q9P8C9
22	6	42.9	438	2	Q9KGU2
23	6	42.9	512	10	Q9LVX9
24	6	42.9	528	2	Q9A4N9
25	6	42.9	554	10	P93719
26	6	42.9	629	10	Q39174
27	6	42.9	629	10	Q9FX95
28	6	42.9	655	3	Q06623
29	6	42.9	706	2	Q9A202
30	6	42.9	840	2	Q9EYF9
31	6	42.9	871	2	Q9L2D2
32	6	42.9	1157	1	O26587
33	6	42.9	1258	13	O93520
34	6	42.9	1261	2	O69247
35	6	42.9	1288	2	O66071
36	6	42.9	2037	5	Q9VIS8
37	5	35.7	65	2	Q9KF64
38	5	35.7	65	4	O43616
39	5	35.7	68	2	Q9K893
40	5	35.7	71	2	P94537
41	5	35.7	72	2	Q9AN03
42	5	35.7	81	2	Q9A9R3
43	5	35.7	85	5	Q9BHC3
44	5	35.7	89	12	Q71656
45	5	35.7	94	2	Q9S2R9

ALIGNMENTS

RESULT 1
071028
ID 071028 PRELIMINARY; PRT; 158 AA.
AC 071028;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE DUTPASE (EC 3.6.1.23) (DUTP PYROPHOSPHATASE) (DEOXYURIDINE-
DE TRIPHOSPHATASE).
OS Archaeal virus SIRV.
OC Viruses; unclassified viruses.
OX NCBI_TaxID=66287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98165772; PubMed=9497317;
RA Prangisvilli D., Klenk H.P., Jakobs G., Schmiechen A., Hanselmann C.,
RA Holz I., Zillig W.;
RT "Biochemical and phylogenetic characterization of the dUTPase from the
archaeal virus SIRV."
RL J. Biol. Chem. 273:6024-6029(1998).
CC -1- CATALYTIC ACTIVITY: DUTP + H(2)O = DUMP + PYROPHOSPHATE.
DR EMBL; AF022221; AAC15873.1; -;
DR InterPro; IPR001428; dUTPase.
DR InterPro; IPR003232; dCTP_deaminase.
DR Pfam; PF00692; dUTPase; 1.
KW Hydrolase. PD004900; dCTP_deaminase; 1.
SQ SEQUENCE. 158 AA; 17906 MW; 5BA47765E0889190 CRC64;

Query Match 100.0%; Score 14; DB 12; Length 158;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTIVDAGFEGOLTI 14
Db 96 PTIVDAGFEGOLTI 109

RESULT 2
Q9RB12 PRELIMINARY; PRT; 89 AA.
ID Q9RB12

AC Q9RB12: 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE YTHD.
 GN YTHD.
 OS *Erwinia carotovora* subsp. *carotovora*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCRI193;
 RX MEDLINE=99392457; PubMed=10463161;
 RA Vincent-Sealy L.V., Thomas J.D., Commander P., Salmond G.P.C.;
 RT "Brwinia carotovora Daba mutants: evidence for a periplasmic-stress
 RT signal transduction system affecting transcription of genes encoding
 RT secreted proteins.";
 RL Microbiology 145:1945-1958(1999).
 DR EMBL; AF146615; AAD47615.1; -;
 SQ SEQUENCE 89 AA; 10402 MW; 228FDD71E865481 CRC64;

 Query Match 50.0%; Score 7; DB 2; Length 89;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 6 AGFEGQL 12
 Db 36 AGFEGQL 42
 |||||

 RESULT 3
 Q9G084 PRELIMINARY; PRT; 328 AA.
 ID Q9G084
 AC Q9G084
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE COMPLETE GENOME
 OS Bacteriophage S13
 OC Viruses; ssDNA viruses; Microviridae; Microvirus.
 OX NCBI_TaxID=10844;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ANC;
 RA Wichman H.A., Scott L.A., Yarber C.D., Bull J.J.;
 RT "Experimental Evolution Recapitulates Natural Evolution.";
 RL Proc. R. Soc. Lond., B, Biol. Sci. 0:0-0(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ANC;
 RA Yarber C.D., Wichman H.A., Bull J.J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF274751; AAC29863.1; -;
 SQ SEQUENCE 328 AA; 34377 MW; EB942BA33DD2B4D9 CRC64;

 Query Match 50.0%; Score 7; DB 9; Length 328;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 3 IVDAGFE 9
 Db 143 IVDAGFE 149
 |||||

 RESULT 4
 Q9LK94 PRELIMINARY; PRT; 488 AA.
 ID Q9LK94
 AC Q9LK94
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE CYTOSOLIC MONOHYDROCARBOXYLATE REDUCTASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
 CC CLASS-I.
 DR EMBL; AP000371; BAB02528.1; -;
 DR InterPro; IPR001327; FAD_pyr_redox.
 DR InterPro; IPR001103; Pyridine_redox_2.
 DR InterPro; IPR001100; pyr_redox.
 DR Pfam; PF00070; pyr_redox; 1.
 DR PRINTS; PR00368; FADPNR.
 DR PRINTS; PR00411; PNDRDTASEII.
 DR PRINTS; PR00469; PNDRDTASEII.
 KW FAD: Flavoprotein; Oxidoreductase; Redox-active center.
 SQ SEQUENCE 488 AA; 53526 MW; AC21B11991178DFA CRC64;

 Query Match 50.0%; Score 7; DB 10; Length 488;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 8 FEGOLTI 14
 Db 267 FEGOLTI 273
 |||||

 RESULT 5
 P74723 PRELIMINARY; PRT; 110 AA.
 ID P74723
 AC P74723
 DT 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
 DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
 DE HYPOTHETICAL 11.8 KDA PROTEIN.
 GN SLR0587.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirose M., Sugita M., Sugita M., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90917; BAA18843.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 110 AA; 11829 MW; 320B467B33BF268F CRC64;

Query Match 42.9%; Score 6; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 FEGOLT 13
 DB 85 FEGOLT 90

RESULT 6
 Q49668 PRELIMINARY; PRT; 118 AA.
 AC Q49668;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE RPSB.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96084954; PubMed=7489918;
 RA Doukhan L., Predich M., Nair G., Dussurget O., Mandic-Mulec I.,
 RA Cole S.T., Smith D.R., Smith I.;
 RT "Genomic organization of the mycobacterial sigma gene cluster.";
 RL Gene 163:67-70(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robison K.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U00012; AAA8594.1; -;
 SQ SEQUENCE 118 AA; 12963 MW; F9C9885EFA48FF8 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 118;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 VDAGFE 9
 DB 90 VDAGFE 95

RESULT 7
 O48474 PRELIMINARY; PRT; 184 AA.
 AC O48474;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DE COMPLETE NUCLEOTIDE SEQUENCE.
 OS Bacteriophage SP1.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=10724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Alonso J.C., Luder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X37918; CAA6651.1; -;
 SQ SEQUENCE 184 AA; 20865 MW; AB00A13881303F75 CRC64;

Query Match 42.9%; Score 6; DB 9; Length 184;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 EGQITI 14
 DB 173 EGQITI 178

RESULT 8
 Q9GY57 PRELIMINARY; PRT; 202 AA.
 ID Q9GY57
 AC Q9GY57;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 21.4 KDA PROTEIN (FRAGMENT).
 GN LM12.89.
 OS Leishmania major.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
 RA Oliver K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL390114; CAC01993.1; -;
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 202
 SQ SEQUENCE 202 AA; 21440 MW; 890E3248C4410E5A CRC64;

Query Match 42.9%; Score 6; DB 5; Length 202;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 DAGFEG 10
 DB 130 DAGFEG 135

RESULT 9
 Q22723 PRELIMINARY; PRT; 218 AA.
 ID Q22723
 AC Q22723;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE COSMID T24C12.
 GN T24C12.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RN Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Miller N., Stellyes L.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

```

DR EMBL: U41037; AAA82390.1; -.
SQ SEQUENCE 218 AA; 24918 MW; 353C7801098301E7 CRC64;

Query Match 42.9%; Score 6; DB 5; Length 218;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDA GF 8
Db 50 IVDA GF 55

RESULT 10
ID Q9DBX3 PRELIMINARY; PRT; 218 AA.
AC Q9DBX3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE N-ACETYL TRANSFERASE 2.
GN NAT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H., Ashburner M., Batilov B., Casavant R.,
RA Fleischmann W., Gaasterland T., Giasi C., King B., Koculwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Piro G., Mackenroth J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita T., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh S.,
RA Blake J., Boffelli D., Bolunga N., Carninci P., de Boer M.F.,
RA Gwosteinich S.J., Hill D., Fletcher C., Fujita M., Gariboldi M.,
RA Rusting P., Hofmann M., Hume D.A., Kamiya N., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli S., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez J., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection.;
RL Nature 409:685-690(2001).
DR EMBL: AK007583; BAB25121.1; -.
DR MGD; MGI:109201; Nat2.
DR InterPro; IPR001447; Acetyltransf2.
DR Pfam; PF00797; Acetyltransf2; 1.
DR ProDom; PD002783; Acetyltransf2; 1.
SQ SEQUENCE 218 AA; 25378 MW; 1E3043003F4CFCE CRC64;

Query Match 42.9%; Score 6; DB 11; Length 218;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDA GF 8
Db 120 IVDA GF 125

RESULT 11
ID Q9ABU6 PRELIMINARY; PRT; 260 AA.
AC Q9ABU6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE N-ACETYL TRANSFERASE 2.
GN NAT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H., Ashburner M., Batilov B., Casavant R.,
RA Fleischmann W., Gaasterland T., Giasi C., King B., Koculwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Piro G., Mackenroth J.,
RA Schriml L.M., Straubli F., Suzuki R., Tomita T., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh S.,
RA Blake J., Boffelli D., Bolunga N., Carninci P., de Boer M.F.,
RA Gwosteinich S.J., Hill D., Fletcher C., Fujita M., Gariboldi M.,
RA Rusting P., Hofmann M., Hume D.A., Kamiya N., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli S., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez J., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection.;
RL Nature 409:685-690(2001).
DR EMBL: AK007583; BAB25121.1; -.
DR MGD; MGI:109201; Nat2.
DR InterPro; IPR001447; Acetyltransf2.
DR Pfam; PF00797; Acetyltransf2; 1.
DR ProDom; PD002783; Acetyltransf2; 1.
SQ SEQUENCE 218 AA; 25378 MW; 1E3043003F4CFCE CRC64;

Query Match 42.9%; Score 6; DB 2; Length 282;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDA GF 8
Db 176 IVDA GF 181

RESULT 13
Q9HAQ5

```


ID Q9HAQ5 PRELIMINARY; PRT; 290 AA.
 AC Q9HAQ5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ARYLAMINE N-ACETYLTRANSFERASE 1 (EC 2.3.1.5).
 GN NAT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cascorbi I., Roots I., Brockmoller J.;
 RT "Homo sapiens NAT1 gene for arylamine N-acetyltransferase 1, NAT1*11C
 allele";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF308866; AG23842.1; -;
 DR InterPro; IPR001447; Acetyltransf2.
 DR Pfam; PF00797; Acetyltransf2; 1.
 DR ProDom; PD002783; Acetyltransf2; 1.
 KW Transferase; Acyltransferase.
 SQ SEQUENCE 290 AA; 33882 MW; 7FF417F3C24FDB11 CRC64;

 Query Match 42.9%; Score 6; DB 4; Length 290;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 3 IVDAGF 8
 DB 120 IVDAGF 125

 RESULT 14
 ID Q9HP97 PRELIMINARY; PRT; 325 AA.
 AC Q9HP97;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE VNG1740C.
 GN VNG1740C.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angewine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AE005079; AAG19973.1; -;
 DR InterPro; IPR001934; HMA.
 DR PROSITE; PS01047; HMA; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 325 AA; 33584 MW; CB427B9A9E190A88 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 325;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 4 VDAGFE 9
 DB 120 IVDAGFE 125

Db 146 VDAGFE 151

 RESULT 15
 ID Q9RW33 PRELIMINARY; PRT; 349 AA.
 AC Q9RW33;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN DR0836.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathavan J.J., Lam P., McDonald L., Otterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans RI.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001938; AAF10415.1; -;
 DR TIGR; DR0836; -;
 KW Complete proteome.
 SQ SEQUENCE 349 AA; 38263 MW; 77145B0A11426EAA CRC64;

 Query Match 42.9%; Score 6; DB 2; Length 349;
 Best Local Similarity 100.0%; Pred. No. 91;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 8 FEGQLT 13
 DB 13 FEGQLT 18

Search completed: January 31, 2002, 13:37:45
 Job time: 170 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:20:14 ; Search time 78.64 Seconds
(without alignments)
13.561 Million cell updates/sec

Title: US-08-957-709-76

Perfect score: 14

Sequence: 1 PTIVDAGFEGQLTI 14

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	100.0	173	2 S26382	probable dCTP deam
2	11	78.6	181	2 D72724	probable dCTP deam
3	8	57.1	180	2 A70439	probable dCTP deam
4	7	50.0	328	2 JS0459	gene H protein - p
5	7	50.0	984	2 JN0658	restriction endonu
6	6	42.9	110	2 S76931	hypothetical prote
7	6	42.9	184	2 T42314	hypothetical prote
8	6	42.9	218	2 T34373	hypothetical prote
9	6	42.9	242	2 I78931	arylamine N-acetyl
10	6	42.9	287	1 A28168	arylamine N-acetyl
11	6	42.9	290	1 XYRBP	arylamine N-acetyl
12	6	42.9	290	1 XYRBM	arylamine N-acetyl
13	6	42.9	290	1 XYCHV0	arylamine N-acetyl
14	6	42.9	290	2 A34585	arylamine N-acetyl
15	6	42.9	290	2 B61267	arylamine N-acetyl
16	6	42.9	290	2 I78930	arylamine N-acetyl
17	6	42.9	325	2 A84326	hypothetical prote
18	6	42.9	326	2 B64230	pyruvate dehydroge
19	6	42.9	327	2 B42653	pyruvate dehydroge
20	6	42.9	327	2 S73772	pyruvate dehydroge
21	6	42.9	328	1 ZHBPF4	gene H protein - p
22	6	42.9	349	2 H75468	conserved hypothet
23	6	42.9	357	2 T38405	hypothetical prote
24	6	42.9	386	2 T39461	hypothetical prote
25	6	42.9	406	2 H82397	maltose ABC transp
26	6	42.9	440	2 S48953	hypothetical prote
27	6	42.9	525	2 G75527	lysyl-tRNA synthet
28	6	42.9	578	2 S50446	VAC8 protein - yea
29	6	42.9	587	2 A34084	extracellular prot

ALIGNMENTS

RESULT 1

S26382

Probable dCTP deaminase (EC 3.5.4.13) [similarity] - Desulfurolobus ambivalens
N;Alternate names: Hypothetical protein 3 lig-region

C;Species: Desulfurolobus ambivalens

C;Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 31-Mar-2000

C;Accession: S26382

R;Kletzin, A.

Nucleic Acids Res. 20, 5389-5396, 1992

A;Title: Molecular characterisation of a DNA ligase gene of the extremely thermophilic
A;Reference number: S26382; MUID:93065206

A;Accession: S26382

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-173 <KLE>

A;Cross-references: EMBL:X63438; NID:G40784; PIDN:CAA45033.1; PID:G40785

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1991

C;Superfamily: dCTP deaminase

C;Keywords: hydrolase

Query Match 100.0%; Score 14; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PTIVDAGFEGQLTI 14

|||||
Db 109 PTIVDAGFEGQLTI 122

RESULT 2

D72724

Probable dCTP deaminase (EC 3.5.4.13) APE0333 [similarity] - Aeropyrum pernix (strain
C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C;Accession: D72724

R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A;Reference number: A72450; MUID:99310339

A;Accession: D72724

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-181 <KAW>

A;Cross-references: DDBJ:AP000059; NID:G5103911; PIDN:BAA79288.1; PID:G5103972

A;Experimental source: strain K1

C;Genetics:

A;Gene: APE0333

C;Superfamily: dCTP deaminase

C;Keywords: hydrolase

Query Match 78.6% Score 11; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDAGFEGQLT 14
|||||
Db 117 VDAGFEGQLT 127

RESULT 3
A70439
probable dCTP deaminase (EC 3.5.4.13) dcd [similarity] - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
C:Accession: A70439
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70439
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-180 <AOF>
A:Cross-references: GB:AE000747; NID:g2983944; PIDN:AAC07499.1; PID:g2983951; GB:AE00065
A:Experimental source: Strain VF5
C:Genetics:
A:Gene: dcd
C:Superfamily: dCTP deaminase
C:Keywords: hydrolase

Query Match 57.1% Score 8; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDAGFEGQ 11
|||||
Db 112 VDAGFEGQ 119

RESULT 4
JS0459
gene H protein - phage S13
C:Species: phage S13
A:Note: host Escherichia coli
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 20-Sep-1999
C:Accession: JS0459
R:Lau, P.C.K.; Spencer, J.H.
Gene 40, 273-284, 1985
A:Title: Nucleotide sequence and genome organization of bacteriophage S13 DNA.
A:Reference number: JS0450; MUID:86165869
A:Accession: JS0459
A:Molecule type: DNA
A:Residues: 1-328 <LAU>
A:Cross-references: EMBL:M14428; NID:g216089; PIDN:AAA32592.1; PID:g216101
C:Superfamily: phage phi-X174 gene H protein

Query Match 50.0% Score 7; DB 2; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGFE 9
|||||
Db 143 IVDAGFE 149.

RESULT 5
JN0658
restriction endonuclease (EC 3.1.1.-) - Salmonella typhimurium
C:Species: Salmonella typhimurium

C:Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C:Accession: JN0658
R:Dartois, V.; De Backer, O.; Colson, C.
Gene 127, 105-110, 1993
A:Title: Sequence of the Salmonella typhimurium StyLT1 restriction-modification genes
A:Reference number: JN0657; MUID:93252265
A:Accession: JN0658
A:Molecule type: DNA
A:Residues: 1-984 <DAR>
A:Cross-references: GB:M90544
A:Comment: This enzyme belongs to the type-III restriction-modification system.
C:Comment: This enzyme is involved in DNA unwinding at the cleavage site.
C:Genetics:
A:Gene: res
C:Keywords: hydrolase; restriction modification system

Query Match 50.0% Score 7; DB 2; Length 984;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDAGFEG 10
|||||
Db 315 VDAGFEG 321

RESULT 6
S76931
hypothetical protein slr0587 - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76931
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76931
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-110 <KAN>
A:Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAAL8843.1; PID:g165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synecocystis hypothetical protein slr0587

Query Match 42.9% Score 6; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FEGOLT 13
|||||
Db 85 FEGOLT 90

RESULT 7
T42314
hypothetical protein - phage SPPI
C:Species: phage SPPI
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-May-2000
C:Accession: T42314
R:Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, T.A.
Gene 204, 201-212, 1997
A:Title: The complete nucleotide sequence and functional organization of Bacillus sub
A:Reference number: 222157; MUID:98094274
A:Accession: T42314
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-184 <ALO>
A:Cross-references: EMBL:X97918; PIDN:CAA6521.1

Query Match 42.9%; Score 6; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 EGQUTI 14
|||||
Db 173 EGQUTI 178

RESULT 8
T34373
hypothetical protein T24C12.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34373
R:Miller, N.; Stelljes, L.
submitted to the EMBL data Library, November 1995
A:Description: The sequence of C. elegans cosmid T24C12.
A:Reference number: 221514
A:Accession: T34373
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-218 <ML>
A:Cross-references: EMBL:U41037; PIDN:AAA82390.1; CESP:T24C12.4
C:Genetics:
A:Gene: CESP:T24C12.4
A:Introns: 49/1; 109/2; 168/3

Query Match 42.9%; Score 6; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
|||||
Db 50 IVDAGF 55

RESULT 9
I78931
arylamine N-acetyltransferase (EC 2.3.1.5), polymorphic (slow form) - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jul-2000
C:Accession: I78931; I48197
R:Nagata, K.; Ozawa, S.; Miyata, M.; Shimada, M.; Yamazoe, Y.; Kato, R.
Pharmacogenetics 4, 91-100, 1994
A:Title: Primary structure and molecular basis of polymorphic appearance of an acetyltransferase from golden hamster
A:Reference number: I58424; MUID:94362736
A:Accession: I78931
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-242 <RES>
A:Cross-references: GB:S72007; NID:g565127; PIDN:AAB31918.1; PID:g565128
R:Ferguson, R.J.; Doll, M.A.; Baumstark, B.R.; Hein, D.W.
Gene 140, 247-249, 1994
A:Title: Polymorphic arylamine N-acetyltransferase encoding gene (NAT2) from homozygous
A:Reference number: I48197; MUID:94193006
A:Accession: I48197
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-242 <RES>
A:Cross-references: EMBL:U03467; NID:g425279; PIDN:AAB60523.1; PID:g425280
C:Genetics:
A:Gene: NAT2; AT-B/AT-II
C:Superfamily: arylamine acetyltransferase
C:Keywords: acyltransferase; coenzyme A

Query Match 42.9%; Score 6; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
|||||
Db 120 IVDAGF 125

RESULT 10
A28168
arylamine N-acetyltransferase (EC 2.3.1.5) - chicken
N:Alternate names: arylamine acetylase
C:Species: Gallus gallus (chicken)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 05-May-2000
C:Accession: A28168; A28167
R:Ohsako, S.; Ohtomi, M.; Sakamoto, Y.; Uyemura, K.; Deguchi, T.
J. Biol. Chem. 263, 7534-7538, 1988
A:Title: Arylamine N-acetyltransferase from chicken liver. II. Cloning of cDNA and ex
A:Reference number: A28168; MUID:88227946
A:Accession: A28168
A:Molecule type: mRNA
A:Residues: 1-287 <OHS>
A:Cross-references: GB:J03737; NID:g211136; PIDN:AAA48590.1; PID:g211137
R:Deguchi, T.; Sakamoto, Y.; Sasaki, Y.; Uyemura, K.
J. Biol. Chem. 263, 7528-7533, 1988
A:Title: Arylamine N-acetyltransferase from chicken liver. Monoclonal antibodies, imm
A:Reference number: A28167; MUID:88227945
A:Accession: A28167
A:Molecule type: protein
A:Residues: 116-131; 143-164; 272-279 <DEG>
C:Superfamily: arylamine acetyltransferase
C:Keywords: acyltransferase; coenzyme A

Query Match 42.9%; Score 6; DB 1; Length 287;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
|||||
Db 120 IVDAGF 125

RESULT 11
XYRBPR
arylamine N-acetyltransferase (EC 2.3.1.5), polymorphic - rabbit
N:Alternate names: arylamine acetylase
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 05-May-2000
C:Accession: A39870; B39870; S11559; S04138; A60882; A33181
R:Sasaki, Y.; Ohsako, S.; Deguchi, T.
J. Biol. Chem. 266, 13243-13250, 1991
A:Title: Molecular and genetic analyses of arylamine N-acetyltransferase polymorphism
A:Reference number: A39870; MUID:91302353
A:Accession: A39870
A:Molecule type: DNA
A:Residues: 1-290 <SAL>
A:Cross-references: GB:M90378
A:Accession: B39870
A:Molecule type: mRNA
A:Residues: 1-290 <SA2>
A:Cross-references: GB:M90380
R:Blum, M.; Heim, M.; Meyer, U.A.
Nucleic Acids Res. 18, 5295, 1990
A:Title: Nucleotide sequence of rabbit NAT2 encoding polymorphic liver arylamine N-ac
A:Reference number: S11559; MUID:90384846
A:Accession: S11559
A:Molecule type: DNA
A:Residues: 1-290 <BL2>
A:Cross-references: EMBL:X53767; NID:g1647; PIDN:CAA37786.1; PID:g1648
R:Blum, M.; Grant, D.M.; Demierre, A.; Meyer, U.A.
Nucleic Acids Res. 17, 3589, 1989
A:Title: Nucleotide sequence of a full-length cDNA for arylamine N-acetyltransferase
A:Reference number: S04138; MUID:89263801
A:Accession: S04138
A:Molecule type: mRNA

A:Residues: 1-290 <BLU>
A:Cross-references: EMBL:X14673; NID:gl1437; PIDN:CAA32803.1; PID:g1438
R:Andres, H.H.; Vogel, R.S.; Tarr, G.E.; Johnson, L.; Weber, W.W.
Mol. Pharmacol. 31, 446-456, 1987
A:Title: Purification, physicochemical, and kinetic properties of liver acetyl-CoA:arylamine N-acetyltransferase
A:Reference number: A60882; MUID:87201456
A:Accession: A60882
A:Molecule type: protein
A:Residues: 10-13:19-33:118-125,'K';128-151:154-158,'C';160-162,'D';164-165:167-178:189-191:200-210:212-213:215-216:218-219:221-222:224-225:227-228:230-231:233-234:236-237:239-240:242-243:245-246:248-249:251-252:254-255:257-258:260-261:263-264:266-267:269-270:272-273:275-276:278-279:281-282:284-285:287-288:290-291:293-294:296-297:299-300:302-303:305-306:308-309:311-312:314-315:317-318:320-321:323-324:326-327:329-330:332-333:335-336:338-339:341-342:344-345:347-348:350-351:353-354:356-357:359-360:362-363:365-366:368-369:371-372:374-375:377-378:380-381:383-384:386-387:389-390:392-393:395-396:398-399:401-402:404-405:407-408:410-411:413-414:416-417:419-420:422-423:425-426:428-429:431-432:434-435:437-438:440-441:443-444:446-447:449-450:452-453:455-456:458-459:461-462:464-465:467-468:470-471:473-474:476-477:479-480:482-483:485-486:488-489:491-492:494-495:497-498:500-501:503-504:506-507:509-510:512-513:515-516:518-519:521-522:524-525:527-528:530-531:533-534:536-537:539-540:542-543:545-546:548-549:551-552:554-555:557-558:560-561:563-564:566-567:569-570:572-573:575-576:578-579:581-582:584-585:587-588:590-591:593-594:596-597:599-600:602-603:605-606:608-609:611-612:614-615:617-618:620-621:623-624:626-627:629-630:632-633:635-636:638-639:641-642:644-645:647-648:650-651:653-654:656-657:659-660:662-663:665-666:668-669:671-672:674-675:677-678:680-681:683-684:686-687:689-690:692-693:695-696:698-699:701-702:704-705:707-708:710-711:713-714:716-717:719-720:722-723:725-726:728-729:731-732:734-735:737-738:740-741:743-744:746-747:749-750:752-753:755-756:758-759:761-762:764-765:767-768:770-771:773-774:776-777:779-780:782-783:785-786:788-789:791-792:794-795:797-798:799:801-802:804-805:807-808:810-811:813-814:816-817:819-820:822-823:825-826:828-829:831-832:834-835:837-838:840-841:843-844:846-847:849-850:852-853:855-856:858-859:861-862:864-865:867-868:870-871:873-874:876-877:879-880:882-883:885-886:888-889:891-892:894-895:897-898:899:901-902:904-905:907-908:910-911:913-914:916-917:919-920:922-923:925-926:928-929:931-932:934-935:937-938:940-941:943-944:946-947:949-950:952-953:955-956:958-959:961-962:964-965:967-968:970-971:973-974:976-977:979-980:982-983:985-986:988-989:991-992:994-995:997-998:999:1001-1002:1004-1005:1007-1008:1010-1011:1013-1014:1016-1017:1019-1020:1022-1023:1025-1026:1028-1029:1031-1032:1034-1035:1037-1038:1040-1041:1043-1044:1046-1047:1049-1050:1052-1053:1055-1056:1058-1059:1061-1062:1064-1065:1067-1068:1070-1071:1073-1074:1076-1077:1079-1080:1082-1083:1085-1086:1088-1089:1091-1092:1094-1095:1097-1098:1099:1101-1102:1104-1105:1107-1108:1110-1111:1113-1114:1116-1117:1119-1120:1122-1123:1125-1126:1128-1129:1131-1132:1134-1135:1137-1138:1140-1141:1143-1144:1146-1147:1149-1150:1152-1153:1155-1156:1158-1159:1161-1162:1164-1165:1167-1168:1170-1171:1173-1174:1176-1177:1179-1180:1182-1183:1185-1186:1188-1189:1191-1192:1194-1195:1197-1198:1199:1201-1202:1204-1205:1207-1208:1210-1211:1213-1214:1216-1217:1219-1220:1222-1223:1225-1226:1228-1229:1231-1232:1234-1235:1237-1238:1240-1241:1243-1244:1246-1247:1249-1250:1252-1253:1255-1256:1258-1259:1261-1262:1264-1265:1267-1268:1270-1271:1273-1274:1276-1277:1279-1280:1282-1283:1285-1286:1288-1289:1291-1292:1294-1295:1297-1298:1299:1301-1302:1304-1305:1307-1308:1310-1311:1313-1314:1316-1317:1319-1320:1322-1323:1325-1326:1328-1329:1331-1332:1334-1335:1337-1338:1340-1341:1343-1344:1346-1347:1349-1350:1352-1353:1355-1356:1358-1359:1361-1362:1364-1365:1367-1368:1370-1371:1373-1374:1376-1377:1379-1380:1382-1383:1385-1386:1388-1389:1391-1392:1394-1395:1397-1398:1399:1401-1402:1404-1405:1407-1408:1410-1411:1413-1414:1416-1417:1419-1420:1422-1423:1425-1426:1428-1429:1431-1432:1434-1435:1437-1438:1440-1441:1443-1444:1446-1447:1449-1450:1452-1453:1455-1456:1458-1459:1461-1462:1464-1465:1467-1468:1470-1471:1473-1474:1476-1477:1479-1480:1482-1483:1485-1486:1488-1489:1491-1492:1494-1495:1497-1498:1499:1501-1502:1504-1505:1507-1508:1510-1511:1513-1514:1516-1517:1519-1520:1522-1523:1525-1526:1528-1529:1531-1532:1534-1535:1537-1538:1540-1541:1543-1544:1546-1547:1549-1550:1552-1553:1555-1556:1558-1559:1561-1562:1564-1565:1567-1568:1570-1571:1573-1574:1576-1577:1579-1580:1582-1583:1585-1586:1588-1589:1591-1592:1594-1595:1597-1598:1599:1601-1602:1604-1605:1607-1608:1610-1611:1613-1614:1616-1617:1619-1620:1622-1623:1625-1626:1628-1629:1631-1632:1634-1635:1637-1638:1640-1641:1643-1644:1646-1647:1649-1650:1652-1653:1655-1656:1658-1659:1661-1662:1664-1665:1667-1668:1670-1671:1673-1674:1676-1677:1679-1680:1682-1683:1685-1686:1688-1689:1691-1692:1694-1695:1697-1698:1699:1701-1702:1704-1705:1707-1708:1710-1711:1713-1714:1716-1717:1719-1720:1722-1723:1725-1726:1728-1729:1731-1732:1734-1735:1737-1738:1740-1741:1743-1744:1746-1747:1749-1750:1752-1753:1755-1756:1758-1759:1761-1762:1764-1765:1767-1768:1770-1771:1773-1774:1776-1777:1779-1780:1782-1783:1785-1786:1788-1789:1791-1792:1794-1795:1797-1798:1799:1801-1802:1804-1805:1807-1808:1810-1811:1813-1814:1816-1817:1819-1820:1822-1823:1825-1826:1828-1829:1831-1832:1834-1835:1837-1838:1840-1841:1843-1844:1846-1847:1849-1850:1852-1853:1855-1856:1858-1859:1861-1862:1864-1865:1867-1868:1870-1871:1873-1874:1876-1877:1879-1880:1882-1883:1885-1886:1888-1889:1891-1892:1894-1895:1897-1898:1899:1901-1902:1904-1905:1907-1908:1910-1911:1913-1914:1916-1917:1919-1920:1922-1923:1925-1926:1928-1929:1931-1932:1934-1935:1937-1938:1940-1941:1943-1944:1946-1947:1949-1950:1952-1953:1955-1956:1958-1959:1961-1962:1964-1965:1967-1968:1970-1971:1973-1974:1976-1977:1979-1980:1982-1983:1985-1986:1988-1989:1991-1992:1994-1995:1997-1998:1999:2001-2002:2004-2005:2007-2008:2010-2011:2013-2014:2016-2017:2019-2020:2022-2023:2025-2026:2028-2029:2031-2032:2034-2035:2037-2038:2040-2041:2043-2044:2046-2047:2049-2050:2052-2053:2055-2056:2058-2059:2061-2062:2064-2065:2067-2068:2070-2071:2073-2074:2076-2077:2079-2080:2082-2083:2085-2086:2088-2089:2091-2092:2094-2095:2097-2098:2099:2101-2102:2104-2105:2107-2108:2110-2111:2113-2114:2116-2117:2119-2120:2122-2123:2125-2126:2128-2129:2131-2132:2134-2135:2137-2138:2140-2141:2143-2144:2146-2147:2149-2150:2152-2153:2155-2156:2158-2159:2161-2162:2164-2165:2167-2168:2170-2171:2173-2174:2176-2177:2179-2180:2182-2183:2185-2186:2188-2189:2191-2192:2194-2195:2197-2198:2199:2201-2202:2204-2205:2207-2208:2210-2211:2213-2214:2216-2217:2219-2220:2222-2223:2225-2226:2228-2229:2231-2232:2234-2235:2237-2238:2240-2241:2243-2244:2246-2247:2249-2250:2252-2253:2255-2256:2258-2259:2261-2262:2264-2265:2267-2268:2270-2271:2273-2274:2276-2277:2279-2280:2282-2283:2285-2286:2288-2289:2291-2292:2294-2295:2297-2298:2299:2301-2302:2304-2305:2307-2308:2310-2311:2313-2314:2316-2317:2319-2320:2322-2323:2325-2326:2328-2329:2331-2332:2334-2335:2337-2338:2340-2341:2343-2344:2346-2347:2349-2350:2352-2353:2355-2356:2358-2359:2361-2362:2364-2365:2367-2368:2370-2371:2373-2374:2376-2377:2379-2380:2382-2383:2385-2386:2388-2389:2391-2392:2394-2395:2397-2398:2399:2401-2402:2404-2405:2407-2408:2410-2411:2413-2414:2416-2417:2419-2420:2422-2423:2425-2426:2428-2429:2431-2432:2434-2435:2437-2438:2440-2441:2443-2444:2446-2447:2449-2450:2452-2453:2455-2456:2458-2459:2461-2462:2464-2465:2467-2468:2470-2471:2473-2474:2476-2477:2479-2480:2482-2483:2485-2486:2488-2489:2491-2492:2494-2495:2497-2498:2499:2501-2502:2504-2505:2507-2508:2510-2511:2513-2514:2516-2517:2519-2520:2522-2523:2525-2526:2528-2529:2531-2532:2534-2535:2537-2538:2540-2541:2543-2544:2546-2547:2549-2550:2552-2553:2555-2556:2558-2559:2561-2562:2564-2565:2567-2568:2570-2571:2573-2574:2576-2577:2579-2580:2582-2583:2585-2586:2588-2589:2591-2592:2594-2595:2597-2598:2599:2601-2602:2604-2605:2607-2608:2610-2611:2613-2614:2616-2617:2619-2620:2622-2623:2625-2626:2628-2629:2631-2632:2634-2635:2637-2638:2640-2641:2643-2644:2646-2647:2649-2650:2652-2653:2655-2656:2658-2659:2661-2662:2664-2665:2667-2668:2670-2671:2673-2674:2676-2677:2679-2680:2682-2683:2685-2686:2688-2689:2691-2692:2694-2695:2697-2698:2699:2701-2702:2704-2705:2707-2708:2710-2711:2713-2714:2716-2717:2719-2720:2722-2723:2725-2726:2728-2729:2731-2732:2734-2735:2737-2738:2740-2741:2743-2744:2746-2747:2749-2750:2752-2753:2755-2756:2758-2759:2761-2762:2764-2765:2767-2768:2770-2771:2773-2774:2776-2777:2779-2780:2782-2783:2785-2786:2788-2789:2791-2792:2794-2795:2797-2798:2799:2801-2802:2804-2805:2807-2808:2810-2811:2813-2814:2816-2817:2819-2820:2822-2823:2825-2826:2828-2829:2831-2832:2834-2835:2837-2838:2840-2841:2843-2844:2846-2847:2849-2850:2852-2853:2855-2856:2858-2859:2861-2862:2864-2865:2867-2868:2870-2871:2873-2874:2876-2877:2879-2880:2882-2883:2885-2886:2888-2889:2891-2892:2894-2895:2897-2898:2899:2901-2902:2904-2905:2907-2908:2910-2911:2913-2914:2916-2917:2919-2920:2922-2923:2925-2926:2928-2929:2931-2932:2934-2935:2937-2938:2940-2941:2943-2944:2946-2947:2949-2950:2952-2953:2955-2956:2958-2959:2961-2962:2964-2965:2967-2968:2970-2971:2973-2974:2976-2977:2979-2980:2982-2983:2985-2986:2988-2989:2991-2992:2994-2995:2997-2998:2999:3001-3002:3004-3005:3007-3008:3010-3011:3013-3014:3016-3017:3019-3020:3022-3023:3025-3026:3028-3029:3031-3032:3034-3035:3037-3038:3040-3041:3043-3044:3046-3047:3049-3050:3052-3053:3055-3056:3058-3059:3061-3062:3064-3065:3067-3068:3070-3071:3073-3074:3076-3077:3079-3080:3082-3083:3085-3086:3088-3089:3091-3092:3094-3095:3097-3098:3099:3101-3102:3104-3105:3107-3108:3110-3111:3113-3114:3116-3117:3119-3120:3122-3123:3125-3126:3128-3129:3131-3132:3134-3135:3137-3138:3140-3141:3143-3144:3146-3147:3149-3150:3152-3153:3155-3156:3158-3159:3161-3162:3164-3165:3167-3168:3170-3171:3173-3174:3176-3177:3179-3180:3182-3183:3185-3186:3188-3189:3191-3192:3194-3195:3197-3198:3199:3201-3202:3204-3205:3207-3208:3210-3211:3213-3214:3216-3217:3219-3220:3222-3223:3225-3226:3228-3229:3231-3232:3234-3235:3237-3238:3240-3241:3243-3244:3246-3247:3249-3250:3252-3253:3255-3256:3258-3259:3261-3262:3264-3265:3267-3268:3270-3271:3273-3274:3276-3277:3279-3280:3282-3283:3285-3286:3288-3289:3291-3292:3294-3295:3297-3298:3299:3301-3302:3304-3305:3307-3308:3310-3311:3313-3314:3316-3317:3319-3320:3322-3323:3325-3326:3328-3329:3331-3332:3334-3335:3337-3338:3340-3341:3343-3344:3346-3347:3349-3350:3352-3353:3355-3356:3358-3359:3361-3362:3364-3365:3367-3368:3370-3371:3373-3374:3376-3377:3379-3380:3382-3383:3385-3386:3388-3389:3391-3392:3394-3395:3397-3398:3399:3401-3402:3404-3405:3407-3408:3410-3411:3413-3414:3416-3417:3419-3420:3422-3423:3425-3426:3428-3429:3431-3432:3434-3435:3437-3438:3440-3441:3443-3444:3446-3447:3449-3450:3452-3453:3455-3456:3458-3459:3461-3462:3464-3465:3467-3468:3470-3471:3473-3474:3476-3477:3479-3480:3482-3483:3485-3486:3488-3489:3491-3492:3494-3495:3497-3498:3499:3501-3502:3504-3505:3507-3508:3510-3511:3513-3514:3516-3517:3519-3520:3522-3523:3525-3526:3528-3529:3531-3532:3534-3535:3537-3538:3540-3541:3543-3544:3546-3547:3549-3550:3552-3553:3555-3556:3558-3559:3561-3562:3564-3565:3567-3568:3570-3571:3573-3574:3576-3577:3579-3580:3582-3583:3585-3586:3588-3589:3591-3592:3594-3595:3597-3598:3599:3601-3602:3604-3605:3607-3608:3610-3611:3613-3614:3616-3617:3619-3620:3622-3623:3625-3626:3628-3629:3631-3632:3634-3635:3637-3638:3640-3641:3643-3644:3646-3647:3649-3650:3652-3653:3655-3656:3658-3659:3661-3662:3664-3665:3667-3668:3670-3671:3673-3674:3676-3677:3679-3680:3682-3683:3685-3686:3688-3689:3691-3692:3694-3695:3697-3698:3699:3701-3702:3704-3705:3707-3708:3710-3711:3713-3714:3716-3717:3719-3720:3722-3723:3725-3726:3728-3729:3731-3732:3734-3735:3737-3738:3740-3741:3743-3744:3746-3747:3749-3750:3752-3753:3755-3756:3758-3759:3761-3762:3764-3765:3767-3768:3770-3771:3773-3774:3776-3777:3779-3780:3782-3783:3785-3786:3788-3789:3791-3792:3794-3795:3797-3798:3799:3801-3802:3804-3805:3807-3808:3810-3811:3813-3814:3816-3817:3819-3820:3822-3823:3825-3826:3828-3829:3831-3832:3834-3835:3837-3838:3840-3841:3843-3844:3846-3847:3849-3850:3852-3853:3855-3856:3858-3859:3861-3862:3864-3865:3867-3868:3870-3871:3873-3874:3876-3877:3879-3880:3882-3883:3885-3886:3888-3889:3891-3892:3894-3895:3897-3898:3899:3901-3902:3904-3905:3907-3908:3910-3911:3913-3914:3916-3917:3919-3920:3922-3923:3925-3926:3928-3929:3931-3932

A:Gene: GDB:AAC1; NAT1
A:Cross-references: GDB:125364; OMIM:108345
A:Map position: 8p23.1-8p21.3
C:Function:
A:Description: catalyzes the N-acetylation of arylamines by acetyl-CoA
A:Pathway: detoxification
C:Superfamily: arylamine acetyltransferase
C:Keywords: acyltransferase; coenzyme A; detoxification; liver

Query Match 42.9%; Score 6; DB 2; Length 290;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
| | | | |
Db 120 IVDAGF 125

RESULT 15
B61267
arylamine N-acetyltransferase (EC 2.3.1.5) Nat2 - mouse (strain C57BL/6J)
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 09-Sep-1994 #text_change 05-May-2000
C:Accession: B61267
R:Martell, K.J.; Vatsis, K.P.; Weber, W.W.
Mol. Pharmacol. 40, 218-227, 1991
A:Title: Molecular genetic basis of rapid and slow acetylation in mice.
A:Reference number: A61267; MUID:91342604
A:Accession: B61267
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <MAR>
A:Cross-references: GB:U35887; NID:g1008571; PIDN:AAA78944.1; PID:g1008572
C:Superfamily: arylamine acetyltransferase
C:Keywords: acyltransferase; coenzyme A

Query Match 42.9%; Score 6; DB 2; Length 290;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAGF 8
| | | | |
Db 120 IVDAGF 125

Search completed: January 31, 2002, 13:20:15
Job time: 116 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 13:15:12 ; Search time 65.13 Seconds
(without alignments)
4,837 Million cell updates/sec

Title: US-08-957-709-76
Perfect score: 14
Sequence: 1 PTIVDAGFEGQLTI 14

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 212252 seqs, 22503292 residues

Word size : 0
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCITUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	100.0	14	4	US-08-822-774-57
2	6	42.9	327	3	US-09-154-874-8
3	5	35.7	7	2	US-08-360-606B-2
4	5	35.7	34	6	5514582-27
5	5	35.7	35	4	US-08-810-009-21
6	5	35.7	38	1	US-08-253-155A-45
7	5	35.7	43	2	US-08-533-996A-2
8	5	35.7	62	1	US-08-464-531-117
9	5	35.7	62	2	US-08-461-598-117
10	5	35.7	62	3	US-08-322-137-117
11	5	35.7	97	1	US-08-202-389-15
12	5	35.7	113	6	5248606-47
13	5	35.7	126	1	US-08-202-389-10
14	5	35.7	135	3	US-09-335-409-23
15	5	35.7	219	4	US-08-928-213B-60
16	5	35.7	251	2	US-08-436-748-9
17	5	35.7	251	3	US-08-483-857-9
18	5	35.7	254	1	US-08-378-761A-11
19	5	35.7	254	1	US-08-485-286-11
20	5	35.7	257	1	US-08-378-761A-13
21	5	35.7	257	1	US-08-485-286-13
22	5	35.7	258	4	US-09-097-767A-6
23	5	35.7	261	1	US-08-378-761A-9
24	5	35.7	261	1	US-08-485-286-9
25	5	35.7	262	6	5248606-43
26	5	35.7	265	1	US-08-378-761A-21
27	5	35.7	265	1	US-08-485-286-21

28	5	35.7	271	4	US-09-318-794A-2	Sequence 2, Appli
29	5	35.7	271	4	US-09-318-793A-4	Sequence 4, Appli
30	5	35.7	275	6	5248606-41	Patent No. 5248606
31	5	35.7	276	1	US-08-378-761A-7	Sequence 7, Appli
32	5	35.7	276	1	US-08-485-286-7	Sequence 7, Appli
33	5	35.7	279	4	US-09-097-767A-10	Sequence 10, Appli
34	5	35.7	279	4	US-09-097-767A-14	Sequence 14, Appli
35	5	35.7	285	6	5248606-45	Patent No. 5248606
36	5	35.7	286	1	US-08-378-761A-5	Sequence 5, Appli
37	5	35.7	286	1	US-08-485-286-5	Sequence 5, Appli
38	5	35.7	287	4	US-09-097-767A-17	Sequence 17, Appli
39	5	35.7	290	1	US-08-321-478-7	Sequence 7, Appli
40	5	35.7	290	1	US-08-321-478-8	Sequence 8, Appli
41	5	35.7	290	1	US-08-321-478-9	Sequence 9, Appli
42	5	35.7	290	2	US-08-614-686A-1	Sequence 1, Appli
43	5	35.7	293	4	US-08-980-832-37	Sequence 37, Appli
44	5	35.7	301	1	US-08-378-761A-2	Sequence 2, Appli
45	5	35.7	301	1	US-08-485-286-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-822-774-57
; Sequence 57, Application US/08822774
; Patent No. 6183997
; GENERAL INFORMATION:
; APPLICANT: HOGREFE, Holly
; TITLE OF INVENTION: POLYMERASE ENHANCING FACTOR (PEF)
; TITLE OF INVENTION: Extracts, PEF Protein Complexes, Isolated PEF Proteins,
; TITLE OF INVENTION: and Methods for Purifying and Identifying Same
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David J. Kulik, Evenson, McKeown, Edwards &
; ADDRESSEE: Lenahan, P.L.L.C.
; STREET: 1200 G Street, N.W. Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,774
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: KULIK, David J.
; REGISTRATION NUMBER: 36,576
; REFERENCE/DOCKET NUMBER: 1486/43163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-822-774-57

Query Match 100.0%; Score 14; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 PTIVDAGFEGQLTI 14
|||||

```
Db 1 PTIVDAGFEGOLTI 14

RESULT 2
US-09-154-874-8
; Sequence 8, Application US/09154874
; Patent No. 6054636
; GENERAL INFORMATION:
; APPLICANT: FADER, GARY MICHAEL
; TITLE OF INVENTION: ISOFLAVONE BIOSYNTHETIC ENZYMES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,874
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/931,668
; FILING DATE: SEPTEMBER 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: MAJARTAN, WILLIAM R.
; REGISTRATION NUMBER: 41,173
; REFERENCE/DOCKET NUMBER: BB-1098-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4926
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 8:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-154-874-8

Query Match 42.9%; Score 6; DB 3; Length 327;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VDAGFE 9
Db 300 VDAGFE 305

RESULT 3
US-08-360-606B-2
; Sequence 2, Application US/08360606B
; Patent No. 5919617
; GENERAL INFORMATION:
; APPLICANT: Jnanendra K. Bhattacharjee
; APPLICANT: Richard C. Garrad
; APPLICANT: Paul L. Skatrud
; APPLICANT: Robert P. Peery
; TITLE OF INVENTION: Methods and Reagents for
; TITLE OF INVENTION: Detecting Fungal Pathogens in a
; TITLE OF INVENTION: Biological Sample
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive Suite 3200
; CITY: Chicago
; STATE: Illinois

Query Match 35.7%; Score 5; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFEG 10
Db 25 AGFEG 29

COUNTRY: U.S.A.
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS Word 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/360,606B
FILING DATE: December 21, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Berghoff, Paul H.
REGISTRATION NUMBER: 30,243
REFERENCE/DOCKET NUMBER: 94,319
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acid residues
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEetical: Yes
FRAGMENT TYPE: internal fragment
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
US-08-360-606B-2

Query Match 35.7%; Score 5; DB 2; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DAGFE 9
Db 3 DAGFE 7

RESULT 4
5514582-27
; Patent No. 5514582
; APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.
; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID
; IMMUNOGLOBULINS
; NUMBER OF SEQUENCES: 43
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,670
; FILING DATE: 21-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 986,931
; FILING DATE: 08-DEC-1992
; APPLICATION NUMBER: 808,122
; FILING DATE: 16-DEC-1991
; APPLICATION NUMBER: 440,625
; FILING DATE: 22-NOV-1989
; APPLICATION NUMBER: 315,015
; FILING DATE: 23-FEB-1989
; SEQ ID NO: 27;
; LENGTH: 34
5514582-27

Query Match 35.7%; Score 5; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 AGFEG 10
Db 25 AGFEG 29
```

RESULT 5
US-08-810-009-21
; Sequence 21, Application US/08810009
; Patent No. 6211437
; GENERAL INFORMATION:
; APPLICANT: Briggs, Steven P.
; APPLICANT: Johal, Gurmukh S.
; APPLICANT: Gray, John
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CONTROLLING
; TITLE OF INVENTION: CELL DEATH AND DISEASE RESISTANCE IN PLANTS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BELL, SELTZER, PARK & GIBSON
; STREET: P.O. Drawer 34009
; CITY: Charlotte
; STATE: No. 8211437th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 04-MAR-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5718-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-810-009-21

Query Match 35.7%; Score 5; DB 4; Length 35;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIVDA 6
Db 9 TIVDA 13

RESULT 6
US-08-253-155A-45
; Sequence 45, Application US/08253155A
; Patent No. 5691147
; GENERAL INFORMATION:
; APPLICANT: Gyuris, Jeno
; APPLICANT: Draetta, Giulio
; TITLE OF INVENTION: CDK4 Binding Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/253,155A
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-253-155A-45

Query Match 35.7%; Score 5; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GFEGQ 11
Db 3 GFEGQ 7

RESULT 7
US-08-533-996A-2
; Sequence 2, Application US/08533996A
; Patent No. 5866329
; GENERAL INFORMATION:
; APPLICANT: Demetriou M.D., Achilles A.
; APPLICANT: Ljubimova M.D., Julia Y.
; TITLE OF INVENTION: A NOVEL GENE ASSOCIATED WITH LIVER
; TITLE OF INVENTION: NEOPLASTIC DISEASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,996A
; FILING DATE: 27-SEP-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita Esq., Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: P07 34306
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-996A-2

Query Match 35.7%; Score 5; DB 2; Length 43;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GOLTI 14
|
|
|
|
Db 11 GOLTI 15

RESULT 8
US-08-464-531-117
; Sequence 117, Application US/08464531
; Patent No. 5789184
; GENERAL INFORMATION:
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: BROACH, Jim
; APPLICANT: MANFREDI, John
; APPLICANT: KLEIN, Christine
; APPLICANT: MURPHY, Andrew J.
; APPLICANT: PAUL, Jeremy
; APPLICANT: TRUEHEART, Joshua
; TITLE OF INVENTION: YEAST CELLS ENGINEERED TO PRODUCE
; TITLE OF INVENTION: PHEROMONE SYSTEM PROTEIN SURROGATES, AND USES THEREFOR
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,531
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/322,137
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA: US 08/309,313
; FILING DATE: 20-SEP-1994
; PRIOR APPLICATION DATA: US 08/190,328
; FILING DATE: 31-JAN-1994
; APPLICATION NUMBER: US 08/041,431
; FILING DATE: 31-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: FOLWKES-2G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-464-531-117

Query Match 35.7%; Score 5; DB 1; Length 62;

Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IVDAG 7
|
|
|
|
Db 21 IVDAG 25

RESULT 9
US-08-461-598-117
; Sequence 117, Application US/08461598
; Patent No. 5876951
; GENERAL INFORMATION:
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: BROACH, Jim
; APPLICANT: MANFREDI, John
; APPLICANT: KLEIN, Christine
; APPLICANT: MURPHY, Andrew J.
; APPLICANT: PAUL, Jeremy
; APPLICANT: TRUEHEART, Joshua
; TITLE OF INVENTION: YEAST CELLS ENGINEERED TO PRODUCE
; TITLE OF INVENTION: PHEROMONE SYSTEM PROTEIN SURROGATES, AND USES THEREFOR
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,598
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/322,137
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA: US 08/309,313
; FILING DATE: 20-SEP-1994
; PRIOR APPLICATION DATA: US 08/190,328
; FILING DATE: 31-JAN-1994
; APPLICATION NUMBER: US 08/041,431
; FILING DATE: 31-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: FOLWKES-2F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-461-598-117

Query Match 35.7%; Score 5; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAG 7
Db 21 IVDAG 25

RESULT 10

US-08-322-137-117
; Sequence 117, Application US/08322137
; Patent No. 6100042
; GENERAL INFORMATION:
; APPLICANT: FOLWES, Dana M.
; APPLICANT: BROACH, Jim
; APPLICANT: MANFREDI, John
; APPLICANT: KLEIN, Christine
; APPLICANT: MURPHY, Andrew J.
; APPLICANT: PAUL, Jeremy
; APPLICANT: TRUEHEART, Joshua
; TITLE OF INVENTION: YEAST CELLS ENGINEERED TO PRODUCE
; TITLE OF INVENTION: PHEROMONE SYSTEM PROTEIN SURROGATES, AND USES THEREFOR
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NETMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,137
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/309,313
; FILING DATE: 20-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/190,328
; FILING DATE: 31-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/041,431
; FILING DATE: 31-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, Iver P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: FOLWES-2C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 117:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-322-137-117

Query Match 35.7%; Score 5; DB 3; Length 62;
Best Local Similarity 100.0%; Pred. No. 63;
Matches .5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAG 7
Db 21 IVDAG 25

RESULT 11

US-08-202-389-15
; Sequence 15, Application US/08202389
; Patent No. 5536636
; GENERAL INFORMATION:
; APPLICANT: Freeman Jr., Robert M.
; APPLICANT: Plutzky, Jorge
; APPLICANT: Neel, Benjamin G.
; APPLICANT: Rosenberg, Robert D.
; TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
; TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,389
; FILING DATE: 28-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/983,926
; FILING DATE: 01-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/829,141
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/721,112
; FILING DATE: 26-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: BIH92-05MA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-202-389-15

Query Match 35.7%; Score 5; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IVDAG 7
Db 84 IVDAG 88

RESULT 12

5248606-47
; Patent No. 5248606
; APPLICANT: WALSH, TERENCE A.; HEY, TIMOTHY D.; MORGAN,
; ALICE E.R.
; TITLE OF INVENTION: DNA ENCODING INACTIVE PRECURSOR AND
; ACTIVE FORMS OF MAIZE RIBOSOME INACTIVATION
; NUMBER OF SEQUENCES: 49
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/535,636
; FILING DATE: 11-JUN-1990
; SEQ ID NO: 47;

; LENGTH: 113
5248606-47

Query Match 35.7%; Score 5; DB 6; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels

QY	4	VDAGF	8
Db	30	VDAGF	34

RESULT 13
US-08-202-389-10
; Sequence 10, Application US/08202389
; Patent No. 5536636
; GENERAL INFORMATION:
; APPLICANT: Freeman Jr., Robert M.
; APPLICANT: Plutsky, Jorge
; APPLICANT: Neel, Benjamin G.
; APPLICANT: Rosenberg, Robert D.
; TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
; PHOSPHATASES HAVING SH2 DOMAINS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA

Query Match 35.7%; Score 5; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5: Conservative 0; Mismatches 0; Indels

Qy	3	IVDAG	7
Dp	84	IVDAG	88

```

RESULT 14
US-09-335-409-23
Sequence 23 Application US/09335409
Serial No. 6101029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligoun, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Leach, Joern
TITLE/INVENTION GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE NUMBER: 305822
FILE REFERENCE: 305822
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 23
LENGTH: 135
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-335-409-23

```

Query Match 35.7%; Score 5; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5: Conservative 0; Mismatches 0; Indels

QY 6 AGFEG 10
Dp 39 AGFEG 43

RESULT 15
US-08-928-213B-60
; Sequence 60, Application US/08928213B
; Patent No. 6238905
; GENERAL INFORMATION:
; APPLICANT: McHenry, Charles S.
; Cull, Millard G.
; TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III
; NUMBER OF SEQUENCES: 195
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,213B
FILING DATE: 12-Sep-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: MacKnight, Kamrin T.
REGISTRATION NUMBER: 38,230
REFERENCE/DOCKET NUMBER: ENZYCO-02550
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-705-8410
TELEFAX: 415-357-8338
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid

```
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-08-928-213B-60

Query Match      35.7%; Score 5; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 FEGOL 12
Db      132 FEGOL 136
```

Search completed: January 31, 2002, 13:15:13
Job time: 94 sec

This Page Blank (uspto)